

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2002, 20:57:14 ; Search time 2705 Seconds  
(without alignments)  
247.455 Million cell updates/sec

Title: US-09-700-906A-3  
Perfect score: 23  
Sequence: 1 accaggcgtctcgtggccacat 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 3194402

Minimum DB seq length: 66  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb.ba:\*

2: gb.htg:\*

3: gb.in:\*

4: gb.om:\*

5: gb.ov:\*

6: gb.pat:\*

7: gb.ph:\*

8: gb.pl:\*

9: gb.pr:\*

10: gb.ro:\*

11: gb.sts:\*

12: gb.sy:\*

13: gb.un:\*

14: gb.vi:\*

15: em.ba:\*

16: em.fun:\*

17: em.hum:\*

18: em.in:\*

19: em.mu:\*

20: em.om:\*

21: em.or:\*

22: em.ov:\*

23: em.pat:\*

24: em.ph:\*

25: em.pl:\*

26: em.ro:\*

27: em.sts:\*

28: em.un:\*

29: em.vi:\*

30: em.htg.hum:\*

31: em.htg.inv:\*

32: em.htg.other:\*

33: em.htg.mus:\*

34: em.htg.pln:\*

35: em.htg.rod:\*

36: em.htg.mam:\*

37: em.htg.vrt:\*

38: em.sy:\*

39: em.htgo.hum:\*

40: em.htgo.mus:\*

41: em.htgo.other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	23	100.0	11435	9	HSWKI67A	X65551 H.sapiens m
C 2	23	100.0	12493	6	AX009576	AX009576 Sequence
C 3	23	100.0	12515	6	AX336950	AX336950 Sequence
C 4	23	100.0	12515	9	HSWKI67	X65550 H.sapiens m
C 5	23	100.0	14041	9	HSKI67	X94762 H.sapiens d
C 6	23	100.0	29965	9	HSAMAK67	X74107 H.sapiens g
C 7	23	100.0	14046	9	AL355529	AL355529 Human DNA
C 8	18.8	81.7	15195	2	AC126512	AC126512 Rattus no
C 9	18.2	79.1	14563	9	AL512622	AL512622 Human DNA
C 10	17.8	77.4	13585	9	AC091714	AC091714 Papio cyn
C 11	17.8	77.4	16452	2	AC127631	AC127631 Rattus no
C 12	17.8	77.4	16823	9	AC007663	AC007663 Homo sapi
C 13	17.8	77.4	17484	9	AC006549	AC006549 Homo sapi
C 14	17.8	77.4	18429	9	AC058790	AC058790 Homo sapi
C 15	17.2	74.8	39482	3	CEFO4A8	Z35663 Caenorhabdi
C 16	17.2	74.8	79660	3	AC004249	AC004249 Drosophil
C 17	17.2	74.8	10918	2	AC127160	AC127160 Rattus no
C 18	17.2	74.8	12550	2	AC092732	AC092732 Felis cat
C 19	17.2	74.8	12756	2	AC100788	AC100788 Homo sapi
C 20	17.2	74.8	14345	2	AC017124	AC017124 Drosophil
C 21	17.2	74.8	15407	2	HS1164K7	AL592209 Homo sapi
C 22	17.2	74.8	15433	9	AC117409	AC117409 Homo sapi
C 23	17.2	74.8	17824	9	AC008151	AC008151 Homo sapi
C 24	17.2	74.8	18542	3	AC099027	AC099027 Drosophil
C 25	17.2	74.8	19491	2	AC122284	AC122284 Mus muscu
C 26	17.2	74.8	20242	2	AC130160	AC130160 Rattus no
C 27	17.2	74.8	20387	2	AC040970	AC040970 Homo sapi
C 28	17.2	74.8	20520	2	AC122005	AC122005 Mus muscu
C 29	17.2	74.8	22429	2	AC118218	AC118218 Mus muscu
C 30	17.2	74.8	25822	3	AE003807	AE003807 Drosophil
C 31	17.2	74.8	30251	2	AC076961	AC076961 Homo sapi
C 32	17.2	74.8	34484	2	AC092902	AC092902 Homo sapi
C 33	17	73.9	12654	2	AC026156	AC026156 Homo sapi
C 34	17	73.9	13806	9	AC092473	AC092473 Homo sapi
C 35	16.8	73.0	1557	6	AX431834	AX431834 Sequence
C 36	16.8	73.0	2054	9	HSAM803725	HSAM803725 Homo sapi
C 37	16.8	73.0	2279	9	HSWKI67A	HSWKI67A Sequence
C 38	16.8	73.0	4042	6	ARI63474	ARI63474 Sequence
C 39	16.8	73.0	4592	9	D83195	D83195 Human DNASE
C 40	16.8	73.0	41308	9	AC005203	AC005203 Homo sapi
C 41	16.8	73.0	11483	2	OSJN00036	AL606598 Oryza sat
C 42	16.8	73.0	15756	9	AC022517	AC022517 Homo sapi
C 43	16.8	73.0	16519	2	AC121309	AC121309 Mus muscu
C 44	16.8	73.0	16935	9	AL158165	AL158165 Human DNA
C 45	16.8	73.0	18168	2	AC021618	AC021618 Homo sapi

ALIGNMENTS

RESULT 1  
HSWKI67A/c  
LOCUS  
DEFINITION H.sapiens mki67a mRNA (short type) for antigen of monoclonal antibody Ki-67.  
ACCESSION X65551  
VERSION X65551.1 GI:415820  
KEYWORDS antigen; monoclonal antibody.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 11435)  
AUTHORS Gerdes Fors,J.  
TITLE Direct Submission

JOURNAL

Submitted (11-APR-1992) J. Gerdes, Forschungsinstitut Borstel, Div. Molecular Immunology, Parkallee 22, 2061 Borstel, FRG

REMARK

sequence revised by author 14-JUL-93 and 08-OCT-93

REFERENCE

2 (bases 1 to 11435)

AUTHORS

Schluter, C., Duchro, M., Wohlenberg, C., Becker, M.H., Key, G.,

TITLE

Flad, H.D. and Gerdes, J.  
The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiquitous nuclear protein with numerous repeated elements, representing a new kind of cell cycle-maintaining proteins

JOURNAL

Cell Biol. 123 (3), 513-522 (1993)

MEDLINE

94043435

PUBMED

8227122

FEATURES

source

Location/Qualifiers

1..11435

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="10q25-qter"

/cell\_line="IM9"

/clone\_lib="lambda gtl1"

1..107

/number=1

108..288

/number=2

197..8887

/gene="mki-67"

197..8887

/gene="mki-67"

/codon\_start=1

/product="antigen of the monoclonal antibody Ki-67"

/protein\_id="CAA6520.1"

/db\_xref="GI:415821"

/db\_xref="SWISS-PROT:P46013"

/translation="MWPRRLVLTIKRSGVDGPHRPLSLSTCLFGRGIECDIRIOLPVV  
SKOCHKIIEHQEALHNFSTNPTQVNSVIDEVRLLKHGDVITIDRSFRYENSL  
QNGKSTEFPRKIQEAPRRVRSFSSDDESEGIPLKRRRVTFGLHGLFDEN  
LPPNTPLRKAGAPTKSLVMTTPVLLKIIKEQPSGKQESGEIHVEYKAQSLVI  
SPPAPSKRTVADQRRRSCTAPASSQSTVEYRGGERVATCLOKRVYSIRSQH  
DILQICSKRRSGASEANLIVAKSWADVVKLGAKOTQKVIKHGQPSRMNKRQRPAT  
PKKPVGEVSTFGHANSPTIIIGKAHTKGVHVPARPYVLNFIISNOKRDEKL  
SGIAEMFKTPVKEQPOLSTCHIALSENELLKQFGTDSGEPLLPSTSESGGNV  
FSAQNAAQPSDKSASPLLRQCLIRENGNVAKTPRNKYMTSLTKTSTDETPSKT  
VSTVNRSGTEFRNIQKLPVSKSEETNIEICILKRGOKATLLQORRGEKKEIE  
RPFTYKENIELKENDKEMKAMRSRTWGQCAPMSDLTLKSLPDTLMKDTARGON  
LLOTDHAKAPKSEKIKTKMPCQSLQPEINTPTHTKQQLKASLGKVGKVEELAVG  
KFTSGTETHHTHPAGDGKSIRTFKSPKOILDPAARVTGMKKWPTPKKEAQLSE  
DLAGKELFOTPGPEESMTDEKTKIACKSPPEESVDTPSTKQWPKRSILKAEADVE  
EFALAKLTPSAGKAMLTTPKAGGDERDIKAFMGTPOKLDLAGTLPGSKKQLOTPKE  
KAQLEDLAGKELFOTPGHTEELVAAGKTKIPCDSPQSDPVDTPPTSTKORPKRSIR  
KADVEGELLACNLMPSSAGKAMHTPKPSVGEKDIIIFVGPVQKLDLTENLTGSKRR  
PQTPKEEAQLEDLTGKELFOTPGHTEEAAGKTKMPCSESSPPESADPTSTRQ  
PKTPLEKRDVQKLSALKLTOTSGTETHTDKVPGEDKSNAFRETAQKLDPAASV  
TGSKRHPKTEKAQPLEDLAGKELFOTPVCTDKPTTHEKTKIACRSQDPVDTPTS  
SKPOSKSLRVQDVEEFPALRKTPSAGKAMHTPKPAVSGEKNIIYAFMGTPVQKLDL  
TENLTGSKRRITPKKAQLEDLAGKELFOTRGHTEESMTNDKTAACKSOPDL  
DKNFASRRRLTKSLGKVGKVEELAVGKLTOTSGTETHHTPTGDKSNKAFMES  
KOILDSASLRSKQRLTPKRGSEVPEDLAGFIELFOTPSHTKESMTNEKTKVSYR  
ASQDLPVDTPTSSKPPQKRSURKADTEEFLLFRKQTPSAGKAMHTPKPAVGEKDLN  
TEFLGTPVQKLDLPGSNRLOTRKEKAQLEELTGRELFTOTPCTDNPATDEKTT  
KKILCKSPQSDPATPTNPKORPKRSLLKADVEEFLAFRAKLTFSAGKAMHTPKAAVG  
EEKDINTFVGPVQKLDLPGSKRRPQTPKKAQLEDLAGKELFOTPGHTEES  
MTDDKITEVCSQPPDVPKTPSSKQRLKISLGKVGKVEELVPGKLTOTSGTQT  
HRETAGDKSKTAKFESAKQMLDPANTGTGMWRPRTKKEAQLSEDLAGKELFOTP  
DHTDESTDCKTKIACKSPPEESMDTPSTRRPKTPGKRDIVEELSAUKOLTOT  
HTDVPGEDKGINVRETAQKLDPAASVTGSKRPRTPKGAQPLEDLAGKELFQ  
TPVCTDPTTHEKTKIACRSQDPDVPPTIFPKQSKRSURKADVEEELALRKTP  
SVGKAMDTPKPAGGDKKAMFMTGTPVQKLDLPGSKRWPTPKKAQLEDLAG  
FKELFOTPGTDPKTPDEKTKIACKSPQDPVDTASTKQRPKNRLKADVEEFLAL  
KRRTPSAGKAMDTPKPAVSEKNTINFTVETVQKLDLPGSKRROPOTPKKAQAL  
EDLVGKELFOTPGHTEESMTDDKITEVCSQPPESPKTSRKSQRLKIPLVKDMK  
EEPLAVSKLTHTSGTETHHTPTGDKSIAKFKESPKOILDPAASVTGSRQRLTRK  
EKARALEDLVDFELFASAPGHTEESMTIDKNTKIPCKSPPELDTATSTRKPKTRP  
RKVEEELSAVERLTQSGQSTHTKEPAGSGDEGKILVKORAKKPNPVEEPPRRRP  
RAPKEKAQPLEDLAGTSELSTSGHTQESLTAGKATKIPCESPPELVVDTTAKRHL

RTRVQKVQKEEPSAVKFTQTSGETTDADKEPAGEDKIGKALKESAKQTPAPAASVTG  
SRPRAPRESAQAEIEDLAGFKDPAAGHTEESMTDDTKTKIPCKSSPELDTATSSKR  
RPRTRAKVEKVELLAVGLTQTSGETTHDKEPVGEGKTKAFQPAKRVNDAEDV  
IGSRQPRAPKEKAQPLEDLASFOELSOTPGHTELANGAADSTSAKQTPDSCKPL  
KISRRVLRAKPEVGVDSRDPVKSQKSNLPLPFRKGGKDSVGTGKKRLRC  
MPAPEIVEELPASKQKRVAPRARGKSSPVVIMKRSURTSAKRIEPAEELNSDMKT  
NKEHKLDQSDVPENKGISLRSRRQDKTEAQOITEVFLAERIENRNEKKPMKTSPE  
MDIONPDGARKPIPRDKVTENKRLRSARQNESQPKVAEESGQKSAKVLMOQKQK  
KGEAGNSDMLRSRRTKTSQPAASTLESKVSQVTRTSVKRCAENPKKAEDNVCVKKIT  
TRSHRDSEDI"  
289..367  
/gene="mki-67"  
/number=3  
368..483  
/gene="mki-67"  
/number=4  
484..550  
/gene="mki-67"  
/number=5  
561..596  
/gene="mki-67"  
/number=6  
597..772  
/gene="mki-67"  
/note="exon 7 is partially excluded by splicing"  
/number=8  
773..1085  
/gene="mki-67"  
/number=9  
1086..1204  
/gene="mki-67"  
/number=10  
1205..1385  
/gene="mki-67"  
/number=11  
1386..1532  
/gene="mki-67"  
/number=12  
1533..8377  
/gene="mki-67"  
/number=13  
8378..8821  
/gene="mki-67"  
/number=14  
8822..11413  
/number=15  
8914..8919  
polyA\_signal  
9449..9454  
polyA\_signal  
10157..10162  
polyA\_signal  
11388..11393  
polyA\_site  
11414  
BASE COUNT 3802 a 2828 c 2684 g 2121 t  
ORIGIN

Query Match 100.0%; Score 23; DB 9; Length 11435;

Best Local Similarity 100.0%; Pred. No. 0.66;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCAGCGGTCTCGTGGGCACAT 23

|||||

Db 219 ACCAGCGGTCTCGTGGGCACAT 197

RESULT 2

AX009576/c

LOCUS

AX009576

DEFINITION

Sequence 1 from Patent WO9961607.

ACCESSION

AX009576

VERSION

AX009576.1

KEYWORDS

human.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

12493 bp DNA linear

PAT 06-SEP-2000

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 12493)  
Deinert, I., Boehle, A., Gerdes, J., and Flad, H.D.  
Antisense oligonucleotides for treating proliferating cells  
Patent: WO 99/1607-A 1 02-DEC-1999;  
DEINERT, IRINA (DE); BOEHLE, ANDREAS (DE); GERDES, JOHANNES (DE); FLAD  
HANS DIETER (DE); FORSCHUNGSZENTRUM BORSTEL ZENT (DE)

## FEATURES

Location/Qualifiers  
1..12493

/organism="Homo sapiens"

/db\_xref="taxon:9606"

197..9967

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC07498.1"

/db\_xref="GI:9996815"

/translation="MPTRLRLVTKRSVGDGPHPLSLSTCLFGRGTECDIRIOLPVV

SKQCKVEIHEOEALINFSSTNPTQVNGSVDEPVRKKGVDVITIDRSFRYENESL  
QNGKSTEFPRKIREQEPARKVSRSSPSSQDEKADQSKAYSKITEKVSQNGPOVHIK  
NKGEDTADDSKQVAQGTNVHSEHAGRNRAADPIGDFEKSISVKLVSRVYGL  
KSVPTTQCLDNKSNKSPFWKLYESVKKELDVSKQENLVQYCRKSLQTDYATEKES  
ADQLOGTQLLVSRKSPKSGGSHVAEPASPELDQNGKGRDYESVTPSKAVG  
ASFLYEPKMKTPVOYSSQONSPQKHNDLYTTGRRESVNLGKSGFGAGDKTLTP  
RKLSTRNTPAKVEDADSATKPNLSKTRGSIPTDVEVLPTETEHNEPFLMLIT  
QVKRIQDLSLSEKELGTAGCMCSGLPLGSLVDIINFGDSINSESGIPLKRRVSP  
GHLRPELFDENPPTLRKEAPTRKSLVMTTPVLKIKTEQPPQSGKQSGSE  
IHVEVKAQSLVSPSPRPTVADQRRRSCKTAPASSKSQTEVPKRGGERVATC  
LOKRVISRSOHDILQMICRSGSGASEANLIVAKSWADVVKLGAKOTPKVIKHGPQ  
RSMNKRORRATPKPVEVHSQFSTGHANSPCTIIGKAHTEKVHPAPRYLVNPF  
LSNQKMFEDLSGIAEMFKTPYKEQQLTSTCHIAISNENLKGQFQGTDSGEPL  
LPTSESGGVNFFSAQNAKQPSDKSASPLPRQCIRENGVAKTPRNTKTKQKATLL  
QORREGMKTEIERFETFKENIELKENDKMKMRKTWGQKCAPMSDLTDLKSLPD  
TEMLKDTARGONLLOTOHAKAPKSEKGIKMPQOSLOPEIINTPTHTKQOLKASLG  
KVGVEKELLAVGFTTRSGTHTHREPAGDKSIRTFKESPKQLDPAARVTKMKW  
PRTPEKAAQSLDLGAKELFQTPGPSEESWDEKTKIACKSPPPSVDTPTSTKQW  
PKRSRLRAQEEFLALRLKTPKAGDEKDIKAPMGTPVDKTLQDLAGTL  
PGSKRQQLTPKEAQALEDLAGKELFQTPGTEELVAAGTKTKIPCDSPQSDPVDTP  
TSTQRKPRKIRKADVEGELLACNLMPKSAGKAMHTPKPSVGEKDIIFVTPVQKL  
DLNLTGSKRRPTPEKAAQLEDLFGKELFQTPGTEELVAAGTKTKIPCDSPSP  
ESADPTSTRQPTLEKRDVQKELSLKLTQTSGETHTDKVPGEDGKINAFRE  
TAKOKLPDASVSGSKRHPKTEKQAQLEDLAGKELFQTPVCTDKTTHETKTKIAC  
ASQDPDVTPTSSKQPKSRKLVVDVEEFALRKRTPSAGKAMHTPKPAVSGKNIT  
RQMPQVOKLDLTPNLSKRRRLQTPKEAQALEDLAGKELFQTPGTEELVAAGTKTK  
AKVAKSQSDPLDNKPSASRRRLKTSGLKGVKEELLAVGKLTQTSGETHTHTPTG  
DGSKMAFMSPKQILDSAAKSLKQLRTPKGSEVPELAGFIELFOTPSHTKES  
MTNEKTKVSVRASQPDVDTPTSSKQPKSRKLTADTEEFALRKRTPSAGKAMHT  
PKPAVGEKDIINTLGTIPVQKLPDGNLPGSNRRRLQTPKEAQALEDLAGKELFQTP  
CTDNPTADEKTKIKLCKSPQSDPADPTNTKQPKRSLKKAADVEEFALRKLTTPSA  
GKAMHTPKAAVEEKDINTFVTPVEKLDLNLPGSKRRPTPEKAAQALEDLAGKEL  
ELFQTPGTEESMTDDKITEVSKSPQDPVKTPTSSKQRLKISLGVKVEEVLPGV  
KLTQTSKGTQTHRETAGDGKSIAKFKESAKOMLDPANYGTMERWPTPEKAAQSLP  
DLAGEKELFQTPGTEESMTDDKITEVSKSPQDPVKTPTSSKQRLKISLGVKVEEVLPGV  
ELSALKOLTOTHTDKVPGEDGKINAFRETAQOKLPDASVSGSKRQPTPKGAQAP  
LEDLAGKELFQTPVCTDKTTHETKTKIACRSPQDPVPTPTIFKQSKSRKADV  
EESLALRKRTPSGKAMHTPKPAVGEKDIINTFVTPVEKLDLNLPGSKRRPTPEKAAQALEDLAGKELFQTPGTEESMTDDKITEVSKSPQDPVKTPTSSKQRLKISLGVKVEEVLPGV  
KKAQALEDLAGKELFQTPGTEESMTDDKITEVSKSPQDPVKTPTSSKQRLKISLGVKVEEVLPGV  
RKADVEEFALRKRTPSAGKAMHTPKPAVGEKDIINTFVTPVEKLDLNLPGSKRR  
QPOTPKEAQALEDLAGKELFQTPGTEESMTDDKITEVSKSPQDPVKTPTSSKQRLKISLGVKVEEVLPGV  
RLKTLPLVDMKEEPLAVSLKLTISGTTQTHRETAGDGKSIAKFKESQKILDPDPAAS  
VTGSRQRLTKKARALEDLAGKELFQTPGTEESMTDDKITEVSKSPQDPVKTPTSSKQRLKISLGVKVEEVLPGV  
ATSTKRRPRPRAPKEAQALEDLAGKELFQTPGTEESMTDDKITEVSKSPQDPVKTPTSSKQRLKISLGVKVEEVLPGV  
NPVEEFSRRPRAPKEAQALEDLAGKELFQTPGTEESMTDDKITEVSKSPQDPVKTPTSSKQRLKISLGVKVEEVLPGV  
VVDPTASTKRLHTRVQVQVKEPSAVKFTQTSGETHTADKEPAGDEKTKIKPKESA  
KOTPAASVGTGSRPRAPRESAQALEDLAGKELFQTPGTEESMTDDKITEVSKSPQDPVKTPTSSKQRLKISLGVKVEEVLPGV  
PELEDATSKRRPRTRAPKEAQALEDLAGKELFQTPGTEESMTDDKITEVSKSPQDPVKTPTSSKQRLKISLGVKVEEVLPGV  
OPAKRNVDAEDVIGSRPRAPKEAQALEDLAGKELFQTPGTEESMTDDKITEVSKSPQDPVKTPTSSKQRLKISLGVKVEEVLPGV  
DGSVGTGSRPRAPKEAQALEDLAGKELFQTPGTEESMTDDKITEVSKSPQDPVKTPTSSKQRLKISLGVKVEEVLPGV  
PABELNSMDTKNEKHLQDSVPENKGISLSRRODKTAEQOITEVFLAEFLIN  
RNEKPKMTSPMDIQPDGARKPIPRDKVTENKRLSRARQNESSQPKVAESGGQ  
KASVLMQNGKGEAGNSMCLRSRRTKSQPAASLTLESKSVQVRVTRKRAENPK  
KADNVGVCKKITRSHRDSEDI"

## CDS

BASE COUNT 4143 a 3048 c 2929 g 2373 t  
ORIGIN  
Query Match 100.0%; Score 23; DB 6; Length 12493;  
Best Local Similarity 100.0%; Pred. No. 0.67;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ACCAGGCGTCTCGTGGGCCACAT 23  
|||||  
Db 219 ACCAGGCGTCTCGTGGGCCACAT 197  
RESULT 3  
AX336950/c AX336950 12515 bp DNA linear PAT 09-JAN-2002  
DEFINITION Sequence 7459 from Patent WO0194629.  
ACCESSION AX336950  
VERSION AX336950.1 GI:18127669  
KEYWORDS human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1  
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,  
Horrigan, S., Soppet, D.R. and Weaver, Z.  
TITLE Cancer gene determination and therapeutic screening using signature  
gene sets  
JOURNAL Patent: WO 0194629-A 7459 13-DEC-2001;  
Avalon Pharmaceuticals (US)  
FEATURES  
Location/Qualifiers  
1..12515  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 4166 a 3048 c 2928 g 2373 t  
ORIGIN  
Query Match 100.0%; Score 23; DB 6; Length 12515;  
Best Local Similarity 100.0%; Pred. No. 0.67;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ACCAGGCGTCTCGTGGGCCACAT 23  
|||||  
Db 219 ACCAGGCGTCTCGTGGGCCACAT 197  
RESULT 4  
HSMK167/c HSMK167 12515 bp mRNA linear PRI 10-FEB-1999  
DEFINITION H. sapiens mki67a mRNA (long type) for antigen of monoclonal  
antibody Ki-67.  
ACCESSION X65550  
VERSION X65550.1 GI:415818  
KEYWORDS antigen; monoclonal antibody.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 12515)  
AUTHORS Gerdes, J.  
TITLE Direct Submission  
JOURNAL Submitted (11-APR-1992) J. Gerdes, Forschungsinstitut Borstel, Div.  
Molecular Immunology, Parkallee 22, 2061 Borstel, FRG  
REMARK sequence revised by author 13-JUL-93 and 08-OCT-93  
REFERENCE 2 (bases 1 to 12515)  
AUTHORS Schluter, C., Duchrow, M., Wohlenberg, C., Becker, M.H., Key, G.,  
Flad, H.D. and Gerdes, J.  
TITLE The cell proliferation-associated antigen of antibody Ki-67: a very  
large, ubiquitous nuclear protein with numerous repeated elements,  
representing a new kind of cell cycle-maintaining proteins  
J. Cell Biol. 123 (3), 513-522 (1993)  
JOURNAL 94043435  
MEDLINE  
PUBMED 8227122

FEATURES		Location/Qualifiers	
source	exon	1. .12515	PAELASNDMKTNKEEHKLQDSVPENKGISLRSRRODKTEAQOITEVFVLAERIEIN
		/organism="Homo sapiens"	RNEKPKMTSPENDIQNPDGARKPIPRDKVTENKRLCLRSARQNESSQPKVAESGGO
exon	exon	/db_xref="taxon:9606"	KSACLVLONOKGKGAGNSDMLSRKSRKTSQPAASTLESKSQVRVTRSVKRCANPK
		/chromosome="10q25-qter"	KAEDNVCVKIITRSHRSEDI"
gene	exon	/cell_line="IM9"	289. .367
		/clone_lib="lambda gt11"	/gene="mk167"
CDS	exon	1. .107	/number=3
		/number=1	/gene="mk167"
exon	exon	108. .288	/number=4
		/number=2	484. .550
gene	exon	197. .9967	/gene="mk167"
		/gene="mk167"	/number=5
CDS	exon	197. .9967	561. .596
		/gene="mk167"	/number=6
exon	exon	/codon_start=1	597. .1676
		/product="antigen of the monoclonal antibody Ki-67"	/gene="mk167"
exon	exon	/protein_id="CAA46519.1"	/note="partialy excluded by splicing"
		/db_xref="GI:415819"	/number=7
exon	exon	/db_xref="SWISS-PROT:P46013"	1677. .1852
		/translation="MWPTRLRLVTKRSGVDPGPHPLSLSTCLFRGIECDIRIOLPVV	/gene="mk167"
exon	exon	SKQCKIEIHEQEAALINFSSTNPQTQVNGSVIDEPVRLKGDVITIDRSFRVENESL	/number=8
		QNGKSTFEPRKREIPARVSRKSFSDPDEKAQSKAYITEGKVSNGFQVHLK	1853. .2165
exon	exon	NVREDSTADSKOSVAGQTTNVHSSHAGRNRAADPIISGDFKEITSVVLVRYGEL	/gene="mk167"
		KSVPDTQLDNSKNESPFVKLYESVKKELDVSKENLVQYCRKSLQTDYATEKES	/number=9
exon	exon	ADLGQGTQLLVSRKSRPKSGGSHVAEPASPEQELDONKGRDVSQVTPSKAVG	2166. .2284
		ASPLYPEAKMTVPVYVSOQONSQKHKKDLYTTGRRSVNLGKSGFKAGDKTLTP	/gene="mk167"
exon	exon	RKLSRNRTAKVEDAADSATKPENLSKTRGSIPTDVEVLTETELHNEPFLTLMT	/number=10
		QVHERKIQDSKPEKLTAGTCGSLPGLSVVDINNFGDSINESGIPLKKRRVSF	/gene="mk167"
exon	exon	GHLRLPDLNPDPNPTPLKRGAEPTKRKSLVMHTPPVLKIIKEQPPSGKESGE	2285. .2456
		IHVEVQAQSLVISPAPSPRKTVPASQDRRSCKTAPASSKSQTEVPRKGRERVATC	/number=11
exon	exon	LOKRVTSRSHQDILQMICSKRRSGASEANLIVAKSWADVVKLCAKOTQTKVIKHGPO	2457. .2612
		RSNNKRRRPATPKPVGEVHSQFSTGHANSPTCTIIGKAHEKVHVHPAPRYLVNF	/gene="mk167"
exon	exon	ISNQMDFKEDLSGIAEMFKTPKEQQLFSTCHIAISNSENLLGKQFQGTDSGEPL	/number=12
		LPTSESGNVFSAQNAKQPKDSCASPLRQCIRENGNLLAKTPNTYKMTSLET	2613. .9457
exon	exon	KTSDETPEKTSVTVNRSGSTFEFRNIQKLPVESKSEETNEIVECILKRGOKATIL	/gene="mk167"
		QORREGMTKEPTFETKYENIELKENDERKAMKRSRTWGCQKCAPMSDLTDLKSLPD	/number=13
exon	exon	TELMKDARGONLLOTODHAKAPKSEKGIKTKMPCSQLOPEINPTHTKQQLKASLG	9458. .9901
		KVGKVELLAVGKTRTSGETHHREPADGDKSIRTFKESPKOILDPAARVGMKKW	/gene="mk167"
exon	exon	PRTPKEAQSLDLAGKELFOTPGPSEESMTDEKTKIACKSPPESDVPTSTKQW	/number=14
		PKRSLRADVEEFLAUKLTPPSAGKAMLTTPKAGDEKDIKAFMGTPVOKLDLGLTL	9902. .12493
exon	exon	PGSKRQLOTPKEKAQLEDLAGKELFOTPGHTEELVAAGTKTKIPCDSPQSDVPTP	/number=15
		TSTQRKRSIRKADVEGELLACNLMPKSAGKAMHTPKPSVEEKDIIIFVGTPOVKL	9994. .9999
exon	exon	DLTENLGSKRRTQPTKEEAQLEDLTFKELFOTPGHTEEAAGAATTKTMPCESSPP	polyA_signal
		ESADPTSTRRQPTPLEKRDVOKELSALKLTOTSGETHTDKVPGEDKSINARPE	polyA_signal
exon	exon	TAKOKLDPAASVTGSKRHPKTKKAQPLEDLAGKELFOTPVCTDKPTTHEKTKIAC	polyA_signal
		RSQDPDVTPTSSKPSQSKSLRVVDVEEFLAKRRTPSAGKAMHTPKPAVSGKNY	polyA_signal
exon	exon	AFMGTPOKLDLTENLGSKRRLQTPKEKAQLEDLAGKELFOTRGHTEESMTNDKT	polyA_site
		AKVACKSQPDLDKNPASRRRLKTSIGVGVKEELIAGVGLTQTSGETTHTEPTG	12494
exon	exon	DGSKMAFMESPQKQILDASAASLTGSKRQLRTPKGSVPEDLAGFIELFQTPSHTKES	BASE COUNT
		MTNEKTTKVSVASQDPLDVTPTSSKPSQSKSLRKADTEEEFLAFKRTQPSAGKAMHT	4166 a 3048 c 2928 g 2373 t
exon	exon	PKPAVGEKDLNTFLGTPVQKLOPGLPGSNRLQTRKEKAQLELTFGRELFTQTP	ORIGIN
		CTMDPTADEKTKILCKSPQDADPTTNTKQPKRSLRKADVEEFLAFKRLTPSA	Query Match 100.0%; Score 23; DB 9; Length 12515;
exon	exon	GKAMHTPAKVAAGEKDINTFVGTPEVKLDLNLPGSKRRPQTPKEKAQLEDLAGF	Best Local Similarity 100.0%; Pred. No. 0.67;
		ELFQTPGTHTEESMTDDKITEVSKSQPDVPKPTPTSSKQRLKISLGVGVKEEVLPGV	Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
exon	exon	KLQTSQKTTQTHRTADGKSIAKFEKAKOMLDPAHYGTGMERWPTPEKAEQSL	Oy 1 ACCAGGCGTCTCGTGGGCCACAT 23
		DLAGKELFOTPDHTEESTDDKTKIACKSPPEESMDTPTSTRRPKTPLGKRDIYE	
exon	exon	ELSAKALQTTHTDTPGDEDKGINVRETAKOKLDPAASVTGSKRQTPPKGAQVP	Db 219 ACCAGGCGTCTCGTGGGCCACAT 197
		LEDUAGLKELFQTPVCDDKTKIACRSQPDVPVGTPTTIFKPSRSLRKADY	
exon	exon	EEESLALRLTPSVGKAMDTPKAGGDEKDMKFMGTPVOKLDLPGLSKRWQTPP	RESULT 5
		KEKAQLEDLAGKELFOTPDGKTDEKTKIACKSQPDVPDPTASTKQPKRNL	HSK167/c
exon	exon	RKADVEEFLAUKLTPPSAGKAMDTPKPAVSDKNIINTEVTPVOKLDLNLPGSKR	LOCUS HSK167 14041 bp DNA linear PRI 26-JUN-1997
		RLKILPVKDMKEPLAVSLKTRISGETTQTHTEPGDSKSIAKFEKSPKQILDPAAS	DEFINITION H.sapiens DNA for Ki-67 antigen 5'-region (exon 1 & 2).
exon	exon	VTSKQRLKTRKEKRALEDLDLPKELFSPGHTSEESMTIDKNTKIPCKSPPELTP	ACCESSION X94762
		ATSTKPKCTPKRKEEELSAVERLTQTSQSTHTHKEPAGSGDEGILVKLQKRAKLT	VERSION X94762.1 GI:1944550
exon	exon	NPVEEESRRRPAPKEKAQPLEDLAGTSELSTSGHTQESLTAGKATKIPCESPPLE	KEYWORDS Ki-67 gene; monoclonal antibody.
		VVDPTASTKRLRLRVQKVOVEKPSAVKFTQTSGETTADKPEAGDKGTKALKESA	SOURCE Homo sapiens.
exon	exon	KOTPAASVTGSRPRAPRESAQAIEDLAGPKDPAAGHTSEESMTDDKTKIPCKSS	ORGANISM Homo sapiens
		PELEDTATSSKRRPRTAQKVEKELLAVGKLQTSGETTHTEKEPVGEGKGTAKK	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
exon	exon	QPAKNVDAEDIGSRQAPKPEKAQPLEDLASQFELSQTGHTDESLAPLPPFRGGGK	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
		APKQTPDSGKPLCKIRVRLAPKVPVGVVSTRDPVKSQSKSNSTSLPPLPFRGGGK	REFERENCE 1 (bases 1 to 14041)
exon	exon	DGSVTGKRLRCLMCAPEIIEVELPASKQKRVAPRARGKSPEPVIMKRSRLTSKRRIE	AUTHORS Gerdes,J.



```

TITLE      Direct Submission
JOURNAL    Submitted (08-JAN-1996) J. Gerdes, Molecular Immunology,
REFERENCE  2 (bases 1 to 14041)
AUTHORS    Forschunzentrums Borstel, Parkallee 22, D- 23845, Borstel, FRG
TITLE      Gerdes, J.
JOURNAL    Sequence of the human Ki-67 protein gene 5' and promoter region
COMMENT    Unpublished
FEATURES   On Apr 18, 1997 this sequence version replaced gi:1869800.
           Location/Qualifiers
           1..14041
             misc_feature
             1..729
               note="CpG rich island"
             73..180
               exon
               /product="antigen of the monoclonal antibody Ki-67"
               /number=1
               /evidence=experimental
             181..522
               intron
               /number=1
               /evidence=experimental
             523..703
               exon
               /product="antigen of the monoclonal antibody Ki-67"
               /number=2
               /evidence=experimental
             545..568
               repeat_unit
               /note="box 1"
             569..592
               repeat_unit
               /note="box 2"
             704..3100
               intron
               /number=2
               /evidence=experimental
             3101..3179
               exon
               /product="antigen of the monoclonal antibody Ki-67"
               /number=3
               /evidence=experimental
             3180..3273
               intron
               /number=3
               /evidence=experimental
             3274..3389
               exon
               /product="antigen of the monoclonal antibody Ki-67"
               /number=4
               /evidence=experimental
             3390..6958
               intron
               /number=4
               /evidence=experimental
             6959..7025
               exon
               /product="antigen of the monoclonal antibody Ki-67"
               /number=5
               /evidence=experimental
             7026..9741
               intron
               /number=5
               /evidence=experimental
             9742..9787
               exon
               /product="antigen of the monoclonal antibody Ki-67"
               /number=6
               /evidence=experimental
             9788..10270
               intron
               /number=6
               /evidence=experimental
             10271..11350
               exon
               /product="antigen of the monoclonal antibody Ki-67"
               /number=7
               /evidence=experimental
             11351..12673
               intron
               /number=7
               /evidence=experimental
             12674..12849
               exon
               /product="antigen of the monoclonal antibody Ki-67"
               /number=8
               /evidence=experimental
             12850..13825
               intron
               /number=8
               /evidence=experimental
             13826..14138
               exon
               /product="antigen of the monoclonal antibody Ki-67"
BASE COUNT 3477 a 3189 c 3327 g 4048 t
ORIGIN
Query Match 100.0%; Score 23; DB 9; Length 14041;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACCAGGCGTCTCGTGGCCACAT 23
|||||
Db 12357 ACCAGGCGTCTCGTGGCCACAT 12335

RESULT 6
LOCUS      HSAMAK67/c
DEFINITION H.sapiens gene for antigen of the monoclonal antibody Ki-67.
ACCESSION  X74107
VERSION    X74107.1 GI:1370102
KEYWORDS   antigen; monoclonal antibody; nuclear protein; tandem repeat.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 29965)
AUTHORS    Gerdes, J.
TITLE      Direct Submission
JOURNAL    Submitted (12-JUL-1993) J. Gerdes, Forschungsinstitut Borstel, Div.
           Molecular Immunol., Parkallee 22, 23845 Borstel, FRG
REMARK     sequence revised by author 09-OCT-93
REFERENCE  2 (bases 1 to 29965)
AUTHORS    Duchrow, M., Schluter, C., Wohlenberg, C., Flad, H.D. and Gerdes, J.
TITLE      Molecular characterization of the gene locus of the human cell
           proliferation-associated nuclear protein defined by monoclonal
           antibody Ki-67
JOURNAL    Cell Prolif. 29 (1), 1-12 (1996)
```

```
/number=9
/evidence=experimental
14139..14225
/number=9
/evidence=experimental
14226..14344
/product="antigen of the monoclonal antibody Ki-67"
/number=10
/evidence=experimental
14345..14454
/number=10
/evidence=experimental
14455..14626
/product="antigen of the monoclonal antibody Ki-67"
/number=11
/evidence=experimental
14627..15737
/number=11
/evidence=experimental
15738..15893
/product="antigen of the monoclonal antibody Ki-67"
/number=12
/evidence=experimental
15894..16847
/number=12
/evidence=experimental
16848..23692
/product="antigen of the monoclonal antibody Ki-67"
/number=13
/evidence=experimental
17429..17794
/number=13
/evidence=experimental
17795..18159
/number=13
/evidence=experimental
18160..18586
/number=13
/evidence=experimental
18587..18889
/number=13
/evidence=experimental
18890..19252
/number=13
/evidence=experimental
19253..19618
/number=13
/evidence=experimental
19619..19981
/number=13
/evidence=experimental
19982..20349
/number=13
/evidence=experimental
20350..20717
/number=13
/evidence=experimental
20718..21070
/number=13
/evidence=experimental
21071..21433
/number=13
/evidence=experimental
21434..21799
/number=13
/evidence=experimental
21800..22165
/number=13
/evidence=experimental
22166..22585
/number=13
/evidence=experimental
22586..22882
/number=13
/evidence=experimental
22883..23239
/number=13
/evidence=experimental
23693..24569
/number=13
/evidence=experimental
24570..25013
/number=14
/evidence=experimental
25014..27009
/number=14
/evidence=experimental
27010..29601
```

```
/product="antigen of the monoclonal antibody Ki-67"
/number=15
/evidence=experimental
27102..27107
/number=15
/evidence=experimental
27637..27642
/number=15
/evidence=experimental
28345..28350
/number=15
/evidence=experimental
29576..29581
/number=15
/evidence=experimental
8995 a 6446 c 6578 g 7946 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 23; DB 9; Length 29965;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCAGGCGTCTCGTGGCCACAT 23
|||||
Db 634 ACCAGGCGTCTCGTGGCCACAT 612

RESULT 7
AL355529 Human DNA sequence from clone RP11-85C15 on chromosome 10, complete
LOCUS AL355529 140466 bp DNA linear PRI 15-NOV-2001
DEFINITION sequence.
ACCESSION AL355529
VERSION AL355529.21 GI:16972861
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 140466)
Clark,G
Direct Submission
Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 16, 2001 this sequence version replaced gi:16151408.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-85C15 is from the library RPC1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-85C15 it may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP11-264E18 is at 138467 in this
sequence. The true right end of clone RP11-380J17 is at 2000 in
this sequence.
FEATURES
Location/Qualifiers
1..140466
```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-85C15"
/clone_lib="RPC1-11.1"
33521..33577
/note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
113053
/note="Random repeat. Forced join. Gap size estimated to
be approximately 300bp by restriction digest"
36826 a 30925 c 31089 g 41626 t

misc_feature
unsure

BASE COUNT 36826 a 30925 c 31089 g 41626 t
ORIGIN

Query Match 100.0%; Score 23; DB 9; Length 140466;
Best Local Similarity 100.0%; Pred. No. 0.71;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCAGGCTCTCGTGGGCACAT 23
|||||
Db 19280 ACCAGGCTCTCGTGGGCACAT 19302

RESULT 8
AC126512/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-382H23, *** SEQUENCING IN PROGRESS
ACCESSION AC126512
VERSION AC126512.1 GI:21700471
KEYWORDS HTG; HTGS-PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 151959)
Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Jarlssohn,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sison,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,

```

```

Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,C., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 151959)
Worley,K.C.
Direct Submission
Submitted (06-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 151959)
Worley,K.C.
Direct Submission
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYQM
Center clone name: CH230-382H23
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 8991 bases at least Q40
Consensus quality: 97816 bases at least Q30
Consensus quality: 101398 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 67 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1199: contig of 1199 bp in length
* 1200: gap of unknown length
* 1300: contig of 1881 bp in length
* 2981: gap of unknown length
* 3081: gap of unknown length
* 4128: contig of 1048 bp in length
* 4228: gap of unknown length
* 5624: contig of 1396 bp in length
* 5724: gap of unknown length
* 7082: contig of 1358 bp in length
* 7182: gap of unknown length
* 8470: contig of 1288 bp in length
* 8571: gap of unknown length
* 8571: contig of 1385 bp in length
* 9956: gap of unknown length
* 10056: gap of unknown length
* 11544: contig of 1488 bp in length
* 11544: gap of unknown length
* 12709: contig of 1066 bp in length
* 12710: gap of unknown length
* 12809: gap of unknown length
* 13898: contig of 1089 bp in length
* 13998: gap of unknown length
* 15062: contig of 1064 bp in length
* 15162: gap of unknown length
* 15163: contig of 1730 bp in length
* 16893: gap of unknown length
* 16992: gap of unknown length
* 18023: contig of 1031 bp in length
* 18024: gap of unknown length
* 18123: gap of unknown length
* 19259: contig of 1136 bp in length
* 19260: gap of unknown length

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

\* 19360 20594: contig of 1235 bp in length  
\* 20595 20694: gap of unknown length  
\* 20695 22525: contig of 1831 bp in length  
\* 22526 22625: gap of unknown length  
\* 22626 23680: contig of 1055 bp in length  
\* 23681 23780: gap of unknown length  
\* 23781 24853: contig of 1073 bp in length  
\* 24854 24953: gap of unknown length  
\* 24954 27068: contig of 2115 bp in length  
\* 27069 27168: gap of unknown length  
\* 27169 28442: contig of 1274 bp in length  
\* 28443 28542: gap of unknown length  
\* 28543 29589: contig of 1047 bp in length  
\* 29590 29689: gap of unknown length  
\* 29690 30748: contig of 1059 bp in length  
\* 30749 30848: gap of unknown length  
\* 30849 32288: contig of 1440 bp in length  
\* 32289 32388: gap of unknown length  
\* 32389 34352: contig of 1964 bp in length  
\* 34353 34452: gap of unknown length  
\* 34453 36123: contig of 1671 bp in length  
\* 36124 36223: gap of unknown length  
\* 36224 37388: contig of 1160 bp in length  
\* 37389 37488: gap of unknown length  
\* 37489 38856: contig of 1373 bp in length  
\* 38857 38956: gap of unknown length  
\* 38957 41272: contig of 2316 bp in length  
\* 41273 41372: gap of unknown length  
\* 41373 42961: contig of 1589 bp in length  
\* 42962 43061: gap of unknown length  
\* 43062 45423: contig of 2362 bp in length  
\* 45424 45523: gap of unknown length  
\* 45524 47667: contig of 2144 bp in length  
\* 47668 47767: gap of unknown length  
\* 47768 49669: contig of 2202 bp in length  
\* 49670 50069: gap of unknown length  
\* 50070 52171: contig of 2102 bp in length  
\* 52172 52271: gap of unknown length  
\* 52272 54105: contig of 1834 bp in length  
\* 54106 54205: gap of unknown length  
\* 54206 56182: contig of 1977 bp in length  
\* 56183 56282: gap of unknown length  
\* 56283 57715: contig of 1433 bp in length  
\* 57716 57815: gap of unknown length  
\* 57816 60438: contig of 2623 bp in length  
\* 60439 60539: gap of unknown length  
\* 60539 61907: contig of 1368 bp in length  
\* 61907 62006: gap of unknown length  
\* 62007 63893: contig of 1887 bp in length  
\* 63894 63993: gap of unknown length  
\* 63994 65197: contig of 1204 bp in length  
\* 65198 65297: gap of unknown length  
\* 65298 67148: contig of 1851 bp in length  
\* 67149 67248: gap of unknown length  
\* 67249 68927: contig of 1679 bp in length  
\* 68928 69027: gap of unknown length  
\* 69028 71364: contig of 2337 bp in length  
\* 71365 71464: gap of unknown length  
\* 71465 72951: contig of 1487 bp in length  
\* 72952 73051: gap of unknown length  
\* 73052 76101: contig of 3050 bp in length  
\* 76102 76201: gap of unknown length  
\* 76202 78357: contig of 2156 bp in length  
\* 78358 78457: gap of unknown length  
\* 78458 80958: contig of 2501 bp in length  
\* 80959 81058: gap of unknown length  
\* 81059 83003: contig of 1945 bp in length  
\* 83004 83103: gap of unknown length  
\* 83104 85100: contig of 1997 bp in length  
\* 85101 85200: gap of unknown length  
\* 85201 88971: contig of 3771 bp in length  
\* 88972 89071: gap of unknown length  
\* 89072 91097: contig of 2026 bp in length

\* 91098 91197: gap of unknown length  
\* 91198 93142: contig of 1945 bp in length  
\* 93143 93242: gap of unknown length  
\* 93243 95925: contig of 2683 bp in length  
\* 95926 96025: gap of unknown length  
\* 96026 98981: contig of 2956 bp in length  
\* 98982 99081: gap of unknown length

Query Match 81.7%; Score 18.8; DB 2; Length 151959;  
Best Local Similarity 90.9%; Pred. No. 98;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCAGCGTCTCGTGGCCACAT 23  
!!! !!!!!!!!!!!!!!!!  
Db 16329 CCATCGTCTCGTGGCCGCA 16308

RESULT 9  
AL512622/c  
LOCUS Human DNA sequence from clone RP11-140A10 on chromosome 10,  
DEFINITION complete sequence.  
ACCESSION AL512622 AC024653  
VERSION AL512622.9 GI:18369618  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 145630)  
Tracey, A.  
Direct Submission  
Submitted (23-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerv@sanger.ac.uk  
On Jan 24, 2002 this sequence version replaced gi:15020956.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em:, EMBL; Sw:.,  
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep/ This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 10, constructed by the Sanger Centre Chromosome 10  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr10  
RP11-140A10 is from the library RPCI-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone  
RP11-140A10 it may be shorter because we sequence overlapping  
sections only once, except for a short overlap.  
The true right end of clone RP11-140A10 is at 145630 in this  
sequence. The true right end of clone RP11-45A17 is at 2000 in this  
sequence.

FEATURES  
Location/Qualifiers  
1..145630  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"



Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Delavia,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
 Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kratochvil,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,  
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rivas,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Savery,G.,  
 Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,  
 Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs.R.  
 Direct Submission  
 2 (bases 1 to 164523)  
 Worley,K.C.  
 Direct Submission  
 Submitted (18-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: GXY  
 Center clone name: CH230-253F12  
 ----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap: version 0.990329  
 Consensus quality: 123085 bases at least Q40  
 Consensus quality: 127734 bases at least Q30  
 Consensus quality: 131272 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 62 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 1628: contig of 1628 bp in length  
 \* 1629 1728: gap of unknown length  
 \* 1729 2793: contig of 1065 bp in length  
 \* 2794 2893: gap of unknown length  
 \* 2894 3983: contig of 1090 bp in length  
 \* 3984 4083: gap of unknown length  
 \*  
 \* 4084 5719: contig of 1636 bp in length  
 \* 5720 5819: gap of unknown length  
 \* 5820 7455: contig of 1636 bp in length  
 \* 7456 7555: gap of unknown length  
 \* 7556 8705: contig of 1050 bp in length  
 \* 8706 10239: gap of unknown length  
 \* 10240 10339: gap of unknown length  
 \* 10340 11865: contig of 1526 bp in length  
 \* 11866 11965: gap of unknown length  
 \* 11966 13658: contig of 1691 bp in length  
 \* 13659 13756: gap of unknown length  
 \* 13757 15494: contig of 1738 bp in length  
 \* 15495 17015: contig of 1421 bp in length  
 \* 17016 17115: gap of unknown length  
 \* 17116 19291: contig of 2176 bp in length  
 \* 19292 19391: gap of unknown length  
 \* 19392 20489: contig of 1098 bp in length  
 \* 20490 20589: gap of unknown length  
 \* 20590 21915: contig of 1326 bp in length  
 \* 21916 22015: gap of unknown length  
 \* 22016 23987: contig of 1971 bp in length  
 \* 23988 24086: gap of unknown length  
 \* 24087 25771: contig of 1685 bp in length  
 \* 25772 25871: gap of unknown length  
 \* 25872 27398: contig of 1526 bp in length  
 \* 27399 27497: gap of unknown length  
 \* 27498 28780: contig of 1283 bp in length  
 \* 28781 28880: gap of unknown length  
 \* 28881 30447: contig of 1567 bp in length  
 \* 30448 32254: contig of 1707 bp in length  
 \* 32255 32354: gap of unknown length  
 \* 32355 34807: contig of 2453 bp in length  
 \* 34808 34907: gap of unknown length  
 \* 34908 37189: contig of 2282 bp in length  
 \* 37190 38864: contig of 1575 bp in length  
 \* 38865 38964: gap of unknown length  
 \* 38965 40591: contig of 1627 bp in length  
 \* 40592 40691: gap of unknown length  
 \* 40692 42650: contig of 1959 bp in length  
 \* 42651 42750: gap of unknown length  
 \* 42751 45427: contig of 2677 bp in length  
 \* 45428 45527: gap of unknown length  
 \* 45528 47589: contig of 2062 bp in length  
 \* 47590 47689: gap of unknown length  
 \* 47690 49798: contig of 2109 bp in length  
 \* 49799 49898: gap of unknown length  
 \* 49899 51126: contig of 1228 bp in length  
 \* 51127 51226: gap of unknown length  
 \* 51227 53131: contig of 1905 bp in length  
 \* 53132 53231: gap of unknown length  
 \* 53232 55277: contig of 2046 bp in length  
 \* 55278 57246: contig of 1869 bp in length  
 \* 57247 57346: gap of unknown length  
 \* 57347 59394: contig of 2048 bp in length  
 \* 59395 59494: gap of unknown length  
 \* 59495 61271: contig of 1776 bp in length  
 \* 61272 61370: gap of unknown length  
 \* 61371 62415: contig of 1045 bp in length  
 \* 62416 62515: gap of unknown length  
 \* 62516 64435: contig of 1920 bp in length  
 \* 64436 64535: gap of unknown length  
 \* 64536 67351: contig of 2816 bp in length  
 \* 67352 69468: contig of 2017 bp in length  
 \* 69469 69568: gap of unknown length  
 \* 69569 71372: contig of 1804 bp in length  
 \* 71373 71472: gap of unknown length  
 \* 71473 74784: contig of 3312 bp in length

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

```
* 74785 74884: gap of unknown length
* 74885 77727: contig of 2843 bp in length
* 77728 77827: gap of unknown length
* 77828 81301: contig of 3474 bp in length
* 81302 81401: gap of unknown length
* 81402 84515: contig of 3114 bp in length
* 84516 84615: gap of unknown length
* 84616 86554: contig of 2039 bp in length
* 86555 86754: gap of unknown length
* 86755 89683: contig of 2929 bp in length
* 89684 89783: gap of unknown length
* 89784 92502: contig of 2719 bp in length
* 92503 92603: gap of unknown length
* 92604 95664: contig of 3061 bp in length
* 95665 95763: gap of unknown length
* 95764 98683: contig of 2920 bp in length
* 98684 98783: gap of unknown length
* 98784 103140: contig of 4357 bp in length
* 103141 103240: gap of unknown length
* 103241 106637: contig of 3396 bp in length
* 106638 106737: gap of unknown length
* 106738 109305: contig of 2569 bp in length
* 109306 109405: gap of unknown length
* 109406 112922: contig of 3516 bp in length
* 112923 113022: gap of unknown length
* 113023 117375: contig of 4354 bp in length
* 117376 117475: gap of unknown length
* 117476 123341: contig of 5866 bp in length
* 123342 123441: gap of unknown length
* 123442 127005: contig of 3564 bp in length
* 127006 127105: gap of unknown length
* 127106 130679: contig of 3574 bp in length
* 130680 130779: gap of unknown length
* 130780 135645: contig of 4866 bp in length
* 135646 135745: gap of unknown length

Query Match 77.4%; Score 17.8; DB 2; Length 164523;
Best Local Similarity 90.5%; Pred. No. 3.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCAGGCGTCTCGTGGCCAC 22
||||| ||||| ||||| ||||| |||||
Db 56376 CCAGGAGTCCCGTGGCCAC 56396

RESULT 12
AC007663/3
LOCUS AC007663 168239 bp DNA linear PRI 31-MAY-2000
DEFINITION Homo sapiens chromosome 22q11 clone b444p24, complete sequence.
ACCESSION AC007663
VERSION AC007663.29 GI:7940355
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 168239)
Hu, P., Emanuel, B. and Roe, B.A.
Homo sapiens Chromosome 22q11 BAC Clone b444p24 In The BCRL2-GGT
Region
Unpublished
2 (bases 1 to 168239)
Hu, P., Emanuel, B. and Roe, B.A.
Direct Submission
Submitted (27-MAY-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 168239)
Hu, P., Emanuel, B. and Roe, B.A.
Direct Submission
Submitted (25-NOV-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
```

```
REFERENCE
AUTHORS Hu, P., Emanuel, B. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (26-NOV-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Hu, P., Emanuel, B. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Hu, P., Emanuel, B. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Hu, P., Emanuel, B. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On May 19, 2000 this sequence version replaced gi:7212014.
Because these overlapping clones came from different libraries
there are numerous instances of insertions, deletions, and single
nucleotide polymorphisms in the overlapping regions below.
AC006547(pl58119) 114717 179269 (0) overlaps AC007663(b444p24) 1
64569 (103670) AC000097(p201mi8) 102493 162269 (0) overlap
AC007663(b444p24) 1 59787 (108452); AC006549(p215k21) 32674 174840
(0) overlaps AC007663(b444p24) 1 143072 (25167) AC007663(b444p24)
164871 168239 (0) overlaps AC024070(chk89) 1 3369 (35480).
FEATURES
Location/Qualifiers
source 1..168239
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22q11"
/clone="b444p24"
BASE COUNT 34217 a 50533 c 49818 g 33671 t
ORIGIN

Query Match 77.4%; Score 17.8; DB 9; Length 168239;
Best Local Similarity 90.5%; Pred. No. 3.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCAGGCGTCTCGTGGCCAC 22
||||| ||||| ||||| ||||| |||||
Db 130404 CGAGCGTCCCGTGGCCAC 130384

RESULT 13
AC006549/3
LOCUS AC006549 174840 bp DNA linear PRI 31-MAY-2000
DEFINITION Homo sapiens chromosome 22q11 clone p215k21, complete sequence.
ACCESSION AC006549
VERSION AC006549.28 GI:7212020
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174840)
Hu, P., Chen, F., Dumanski, J. and Roe, B.A.
Homo sapiens Chromosome 22q11 PAC Clone p215k21 Distal To DCCR
Region
Unpublished
2 (bases 1 to 174840)
Hu, P., Chen, F., Dumanski, J. and Roe, B.A.
Direct Submission
Submitted (09-FEB-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
```

```
REFERENCE
AUTHORS      3 (bases 1 to 174840)
TITLE        Hu.P., Chen,F., Dumanski,J. and Roe,B.A.
JOURNAL      Direct Submission
              Submitted (25-NOV-1999) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE
AUTHORS      4 (bases 1 to 174840)
TITLE        Hu.P., Chen,F., Dumanski,J. and Roe,B.A.
JOURNAL      Direct Submission
              Submitted (26-NOV-1999) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE
AUTHORS      5 (bases 1 to 174840)
TITLE        Hu.P., Chen,F., Dumanski,J. and Roe,B.A.
JOURNAL      Direct Submission
              Submitted (09-MAR-2000) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE
AUTHORS      6 (bases 1 to 174840)
TITLE        Hu.P., Chen,F., Dumanski,J. and Roe,B.A.
JOURNAL      Direct Submission
              Submitted (31-MAY-2000) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
COMMENT
On Mar 9, 2000 this sequence version replaced gi:6468071.
Because these overlapping clones came from different libraries
there are numerous instances of insertions, deletions, and single
nucleotide polymorphisms in the overlapping regions below.
AC006547(p158119) 81952 179269 (0) overlaps AC006549(p215k21) 1
97222 (77618) AC000097(p201m18) 69724 162269 (0) overlaps
AC006349(p215k21) 1 92440 (82400) AC006549(p215k21) 32674 174840
(0) overlaps AC007663(b44p24) 1 143072 (25167).
FEATURES
    source
        Location/Qualifiers
            1..174840
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="22q11"
                /clone="p215k21"
BASE COUNT      34355 a 52696 c 52199 g 35590 t
ORIGIN
Query Match      77.4%; Score 17.8; DB 9; Length 174840;
Best Local Similarity 90.5%; Pred. No. 3.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2  CCAGCGCTCTCGTGGGCCACA 22
        | ||||| ||||| |||||
Db 162172 CGAGCGTCCGCGGCCACA 162152

RESULT 14
AC058790/c
LOCUS          184929 bp      DNA      linear      PRI 04-AUG-2000
DEFINITION    Homo sapiens chromosome 22q11 clone b518b9, complete sequence.
ACCESSION     AC058790
VERSION       AC058790.14 GI:9690342
KEYWORDS      HTG.
SOURCE        Homo sapiens.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 184929)
              Hu.P., Wu,H., Yang,L., Morrow,B.E. and Roe,B.A.
              Homo sapiens Chromosome 22q11 BAC Clone b518b9 In BCRL2-GGT Region
              Unpublished
              2 (bases 1 to 184929)
              Hu.P., Wu,H., Yang,L., Morrow,B.E. and Roe,B.A.
              Direct Submission
              Submitted (19-APR-2000) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
              3 (bases 1 to 184929)
              Hu.P., Wu,H., Yang,L., Morrow,B.E. and Roe,B.A.
              Direct Submission
              Submitted (04-AUG-2000) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
              On Aug 4, 2000 this sequence version replaced gi:9665183.
              Location/Qualifiers
                  1..184929
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="22q11"
                      /clone="b518b9"
BASE COUNT      37229 a 55467 c 54778 g 37455 t
ORIGIN
Query Match      77.4%; Score 17.8; DB 9; Length 184929;
Best Local Similarity 90.5%; Pred. No. 3.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2  CCAGCGTCTCTCGTGGGCCACA 22
        | ||||| ||||| |||||
Db 142293 CGAGCGTCCGCGGCCACA 142273

RESULT 15
CET04A8
LOCUS          39482 bp      DNA      linear      INV 11-DEC-2001
DEFINITION    Caenorhabditis elegans cosmid T04A8, complete sequence.
ACCESSION     Z35663
VERSION       Z35663.1 GI:527419
KEYWORDS      HTG; 1,4-Alpha-glucan branching enzyme; 50S ribosomal protein L16;
              Amidophosphoribosyltransferase; C-type lectin; calpain thiol
              protease; DNAJ; emb-5 gene; glucose-6-phosphate 1-dehydrogenase;
              Glutamine phosphoribosylpyrophosphate amidotransferase;
              neurofilament triplet M protein; Non-histone chromosomal protein
              HMG-1.
              Caenorhabditis elegans.
              Caenorhabditis elegans
              Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
              Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
              1
              none.
              Genome sequence of the nematode C. elegans: a platform for
              investigating biology. The C. elegans Sequencing Consortium
              Science 282 (5396), 2012-2018 (1998)
              99069613
              9851916
              The C.elegans Sequencing Consortium.
              2 (bases 1 to 39482)
              Palmer,S.
              Direct Submission
              Submitted (04-AUG-1994) Nematode Sequencing Project, Sanger
              Institute, Hinxton, Cambridge CB10 1SA, England and Department of
              Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
              jesus@sanger.ac.uk or rwenematode.wustl.edu
              Coding sequences below are predicted from computer analysis, using
              predictions from Genefinder (P. Green, U. Washington), and other
              available information.
              Current sequence finishing criteria for the C. elegans genome
              sequencing consortium are that all bases are either sequenced
              unambiguously on both strands, or on a single strand with both a
              dye primer and dye terminator reaction, from distinct subclones.
              Exceptions are indicated by an explicit note.
              IMPORTANT: This sequence is not the entire insert of clone T04A8.
              It may be shorter because we only sequence overlapping sections
              once, or longer because we arrange for a small overlap between
              neighbouring submissions.
              The true left end of clone T04A8 is at 1 in this sequence. The true
              right end of clone T04A8 is at 4879 in
              sequence Z35719.
              The true right end of clone T23F11 is at 5962 in this sequence. The
              start of this sequence (1..96) overlaps with the end of sequence
              Z46343.
              The end of this sequence (39470..39482) overlaps with the start of
```



sequence z35719.

For a graphical representation of this sequence and its analysis see: - <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=T04A8>  
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

#### FEATURES

source

```
1. 39482
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="T04A8"
join(285..515,1092..1310,1365..1508,1567..1749,1985..2182)
/gene="T04A8.1"
CDS
join(285..515,1092..1310,1365..1508,1567..1749,1985..2182)
/gene="T04A8.1"
/note="contains similarity to Pfam domain: PF02118
(C.elegans Srg family integral membrane protein),
Score=589.9, E-value=5.1e-174, N=1"
/codon_start=1
/protein_id="CAA84725.1"
/db_xref="GI:3879340"
/db_xref="SWISS-PROT:P46568"
/translaton="MNSRVVFFANFSEYEDPLPECNEDPNVLSLAMYGMQSSYLIV
GAVLNVIVTVFPHGNSYRDNSFVMLYCADAIVGIYINTAEVIFGRFIYITPLCPPIA
SPVEFTSILTKMYAALYSLGFKTESQJFMSFNMTQVFLMKLKLKQILKPVIL
IIFILPLGVKILLSRVYINPAGAGESVNYKDYFPWANISILHLHFHFLCFLVLI
FVFIATIGLMLKORIKASERSLTIVTMAVOTYFASIOIVFVFVAAATPKIRSVL
LQIVSFYDLSLYFSPITALVMSRQLRKLDFNLKQETQISMYPNSEL"
join(2532..2658,2878..3338,3506..3697)
/gene="T04A8.2"
join(2532..2658,2878..3338,3506..3697)
/gene="T04A8.2"
/note="See T23F11.5, contains similarity to Pfam domain:
PF02118 (C.elegans Srg family integral membrane protein),
Score=325.5, E-value=2e-94, N=1"
/codon_start=1
/protein_id="CAA84735.1"
/db_xref="GI:3879350"
/db_xref="SWISS-PROT:P46569"
/translaton="MPSFRENISNDLTIIFECNSNYDITVEVTKWFLQIAYLIPGGI
LNILLYILFKNSEIYASSFFLIYDTCFVSFMIFLDIGRTLVYFTPLCPPIAP
MEYPLGLIKMLHLSHRACKSLQIOLLVNMRSQVIVPIRYGKMMRPLKYLIL
VFVFPESIDNLLISRVYMOPTEGGIYMEYIKKVASQSRQQLIFITIALLTIVCT
SVIFYLVLPLKRLNRVERLSLGTATISMSFIILVVQVN"
complement(join(5293..5696,5899..6060,6234..6659,
6706..6965,7019..7072,7227..7300))
/gene="T04A8.3"
complement(join(5293..5696,5899..6060,6234..6659,
6706..6965,7019..7072,7227..7300))
/gene="T04A8.3"
/note="contains similarity to Pfam domain: PF00059 (Lectin
c-type domain), Score=9.2, E-value=0.0011, N=1"
/codon_start=1
/protein_id="CAA84726.1"
/db_xref="GI:3879341"
/db_xref="SPREMBL:Q22136"
/translaton="MFIDTVIRVLMLESHDIFKMRNRSTRWKVTNARSFLRHRTMRN
PLSTGDNTPYHINNAEASEYKTTCEAEATLICKSGVQVQFGRCYKITMTMRDRA
EEHCKNQDHTSTIAFHREALPFRWMDYFTRSVIRWMDASKVITNDLIYDVEGNYL
LALDGYKLNPLVAIARPKDETAMVLCEYTPPKTKSESNVLLRRYGEIYYPPLVTSE
SVYMRVTSRRINAEADPLADHNYCTELMKPVFRGGEAQALPTQEFVKKLGTNGFRE
YKRVYQGLWCXYVHDMWNDDAEKKNDMGHLSSTFYIEELKULDEMLIEVYPN
DNNIAYWLGAKRRECEGLSKNFTGYGSKDIHDPCARSRVFQWQVQAQNPPIFVGDG
FDYWAERYEPNHSTDSRCLVQMSGSVWYGNKPRNQINDIYCNVEFKFLCGKEA
TTFAYKKNL"
complement(join(7871..8071,8723..8819,8916..9181,
9228..9359,9489..9585,9638..9714))
/gene="T04A8.4"
complement(join(7871..8071,8723..8819,8916..9181,
9228..9359,9489..9585,9638..9714))
```

```
/gene="T04A8.4"
/codon_start=1
/protein_id="CAA84722.1"
/db_xref="GI:3879337"
/db_xref="SPREMBL:Q22133"
/translaton="MPIGSASDTIRKFNVAQAENEVLKKNPYSDTYKIQAFDTKNYG
RPPGSTKEARGIKAGVHVCREILFLCETIDSNADGEPHKYVVFGLFNISYFSOK
LVGMLIARKYGLVHFEGEMLQRODDEKITITMLMSLQETIRSLTASGDPANCQIIRR
NSEPVAFVPPVSKAEAPRPRISSDVTSSRAKFAAPPNTDSLPGVGRPKKPTPKD
TSKAPIHLIRMRKLKIFMGLCLLCYTNAQVDRDWSFOOMCEFWGEQTYRARNYKVO
IYDLSDTIMDLFN"
join(11864..11952,12576..12981,13028..13288,13524..13833,
13882..13997,14293..14457,14502..14597)
/gene="T04A8.5"
join(11864..11952,12576..12981,13028..13288,13524..13833,
13882..13997,14293..14457,14502..14597)
/gene="T04A8.5"
/note="Amidophosphoribosyltransferase (glutamine
phosphoribosylpyrophosphate amidotransferase), contains
similarity to Pfam domain: PF00156 (Phosphoribosyl
transferase domain), Score=15.0, E-value=0.0005, N=1;
PF00310 (Glutamine amidotransferases class-II),
Score=134.9, E-value=7.1e-46, N=1"
/codon_start=1
/protein_id="CAA84723.1"
/db_xref="GI:3879338"
/db_xref="SPREMBL:Q22134"
/translaton="MCGIFGIVAAGNYEHLNVLNLAANGLAOLHGRTESTGLVGS DGI
TADHVAIIHGLVDRVDITEDNISRMGQSIIGHNRYSYTAGKKSGINCVOFPFVYITA
MGTVIAHNGELVDQAKRKVEHGVSLDSTDSELIQAIAKAIALNVKCYQGEOR
DITRELAVTMSALNMSYSLVMTFDRLYAIRDPFGRNPLCVGTGVYKNGNPEAFIAS
SECFAPANAKLDFEVRGEIVELSTGCIKSWQMKNPNTPLAMCIFEVYFARNDSEI
EGQOVTVRECEGKTMALDDLEADIVGNVPDSSLAAIGVASOGSITVEPLHRNSY
VGRSFIENDEMRONAIKMRFGLKKIHGGRIVLVDDSVIRGNTMRTLVMKLRDAGA
KEHLRIASPPVKPCFWMGINPTIKELIAAEKTIPEICQVSGADSVRLYSDGLVSS
VQKGIERAAFPSPGHCTACLTGKYPVAID"
complement(join(14747..14836,14881..15396,15445..15558,
15604..15807))
/gene="T04A8.6"
complement(join(14747..14836,14881..15396,15445..15558,
15604..15807))
/gene="T04A8.6"
/note="contains similarity to Pfam domain: PF00076 (RNA
recognition motif, (a.k.a. RRM, RBD, or RNP domain)),
Score=57.0, E-value=1.3e-13, N=1
CDNA EST yk222a11.3 comes from this gene
CDNA EST yk432f10.3 comes from this gene
CDNA EST yk432f10.5 comes from this gene
CDNA EST yk497a8.3 comes from this gene
CDNA EST yk497a8.5 comes from this gene"
/codon_start=1
/protein_id="CAA84724.1"
/db_xref="GI:3879339"
/db_xref="SPREMBL:Q22135"
/translaton="MVAQNPTSISSGGRKQKQKQKOLKVIYVVKIKRIPDPFF
EKELLYGFERQGNVLRIRVARNRRTGNHGWYVGFONKVAEATAESMGLYLPQER
LCTVMKDELIPKAMRHGPLLVRPSPSLGIAKKDTIARNNTTGRNDVINAKKRVQNL
KTLKLLQNMGIYDFSVSGDTPRKVVIVDNEQVFIQSDVTPKSTPKIVKAAATPRAT
PKATVQKATPKVTPYTKVKVITKPTKEPPEKQFTPKTRAGKAAATAATPOAKSTL
SNTILKSVAAASAPAAVEKKTLRSRKKSL"
join(15523..16784,16831..17000,17047..17601)
/gene="T04A8.8"
join(15523..16784,16831..17000,17047..17601)
/gene="T04A8.8"
/note="cDNA EST yk339d3.3 comes from this gene
CDNA EST yk339d3.5 comes from this gene"
/codon_start=1
/protein_id="CAA84734.1"
/db_xref="GI:3879349"
/db_xref="SPREMBL:Q22144"
/translaton="MDSDDNAYSDYLGNNRAALQKKRQPVRCESDDDDFDFTSGP
SPMRKKTRETQLFDDDDDCVEKETYSQKVRHEIDDDVEEAYRRSIYKRETVIHA
```



JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 109185)  
Worley, K.C.  
Direct Submission

Submitted (14-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 109185)  
Worley, K.C.  
Direct Submission

Submitted (18-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

## COMMENT

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: GVQX  
Center clone name: CH230-460D14

----- Summary Statistics  
Sequencing vector: Plasmid:

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 68267 bases at least Q40

Consensus quality: 72261 bases at least Q30

Consensus quality: 75701 bases at least Q20

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a "working draft" sequence. It currently  
\* consists of 44 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1516: contig of 1516 bp in length  
1517 1616: gap of unknown length  
1617 3136: contig of 1520 bp in length  
3137 3236: gap of unknown length  
3237 4240: contig of 1004 bp in length  
4241 4340: gap of unknown length  
4341 5868: contig of 1528 bp in length  
5869 5968: gap of unknown length  
5969 7089: contig of 1021 bp in length  
7090 8161: contig of 1072 bp in length  
8162 9308: contig of 1047 bp in length  
9309 9408: gap of unknown length  
9409 10838: contig of 1430 bp in length  
10839 10938: gap of unknown length  
10939 12000: contig of 1062 bp in length  
12001 12101: gap of unknown length  
12101 13178: contig of 1078 bp in length  
13179 13278: gap of unknown length  
13279 14788: contig of 1510 bp in length  
14789 14888: gap of unknown length  
14889 16118: contig of 1230 bp in length  
16119 16218: gap of unknown length  
16219 17834: contig of 1616 bp in length  
17835 17934: gap of unknown length  
17935 19555: contig of 1621 bp in length  
19556 21548: contig of 1893 bp in length  
21549 21648: gap of unknown length  
21649 23993: contig of 2345 bp in length  
23994 24093: gap of unknown length  
24094 25811: contig of 1718 bp in length  
25812 25911: gap of unknown length

\* 25912 27904: contig of 1993 bp in length  
27905 28004: gap of unknown length  
28005 29004: contig of 1900 bp in length  
29005 30004: gap of unknown length  
30005 31676: contig of 1672 bp in length  
31677 34223: contig of 2447 bp in length  
34224 34323: gap of unknown length  
34324 36442: contig of 2119 bp in length  
36443 36542: gap of unknown length  
36543 37991: contig of 1449 bp in length  
37992 38091: gap of unknown length  
38092 40125: contig of 2034 bp in length  
40126 40225: gap of unknown length  
40226 43013: contig of 2788 bp in length  
43014 43113: gap of unknown length  
43114 44938: contig of 1825 bp in length  
44939 45038: gap of unknown length  
45039 47723: contig of 2685 bp in length  
47724 47823: gap of unknown length  
47824 49871: contig of 2048 bp in length  
49872 49971: gap of unknown length  
49972 53210: contig of 3239 bp in length  
53211 53310: gap of unknown length  
53311 54891: contig of 1581 bp in length  
54892 54991: gap of unknown length  
54992 57222: contig of 2231 bp in length  
57223 57322: gap of unknown length  
57323 60116: contig of 2794 bp in length  
60117 60216: gap of unknown length  
60217 62503: contig of 2287 bp in length  
62504 62603: gap of unknown length  
62604 64757: contig of 2154 bp in length  
64758 64857: gap of unknown length  
64858 67679: contig of 2822 bp in length  
67680 67779: gap of unknown length  
67780 69622: contig of 1843 bp in length  
69623 69722: gap of unknown length  
69723 72746: contig of 3024 bp in length  
72747 72846: gap of unknown length  
72847 76550: contig of 3704 bp in length  
76551 76650: gap of unknown length  
76651 79635: contig of 2985 bp in length  
79636 83958: gap of unknown length  
83959 84058: contig of 4223 bp in length  
84059 89150: contig of 5092 bp in length  
89151 89250: gap of unknown length  
89251 94921: contig of 5671 bp in length  
94922 95021: gap of unknown length  
95022 103629: contig of 8608 bp in length  
103630 103729: gap of unknown length  
103730 109185: contig of 5456 bp in length.

FEATURES  
source

Location/Qualifiers  
1..109185  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-460D14"

BASE COUNT 30296 a 20794 c 21559 g 30692 t 5844 others  
ORIGIN

Query Match 74.8%; Score 17.2; DB 2; Length 109185;  
Best Local Similarity 86.4%; Pred No. 6.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACCAGGCTCTCGTGGGCCACA 22  
||||| ||||| |||||  
Db 28829 ACCAGGCTCTCGTGGGCCACA 28850

RESULT 18  
AC092732/c  
LOCUS

125506 bp DNA linear HTG 20-JUN-2002

```

DEFINITION Felis catus clone RP86-400G2, WORKING DRAFT SEQUENCE, 4 ordered
pieces.
ACCESSION AC092732
VERSION AC092732.2 GI:21490168
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE cat.
ORGANISM Felis catus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
1 (bases 1 to 125506)
Ahther,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakeley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
Lee-tin,S.-O., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Marquies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 125506)
Green,E.D.
Direct Submission
Submitted (21-JUL-2001) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 125506)
Green,E.D.
Direct Submission
Submitted (20-JUN-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
On Jun 20, 2002 this sequence version replaced gi:14993724.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: cfo
Center clone name: 400G02

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

```

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 124261 bases at least Q40
Consensus quality: 124846 bases at least Q30
Consensus quality: 125088 bases at least Q20
Insert size: 110000; agarose-ff
Insert size: 125206; sum-of-contigs
Quality coverage: 9.88x in Q20 bases; agarose-ff
Quality coverage: 8.68x in Q20 bases; sum-of-contigs

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

```

1 27057: contig of 27057 bp in length
* 27058 109603: contig of 82446 bp in length
* 27158 109603: contig of 82446 bp in length
* 109604 109703: gap of unknown length
* 109704 118983: contig of 9280 bp in length
* 118984 119083: gap of unknown length
* 119084 125506: contig of 6423 bp in length.

```

#### FEATURES

##### source

```

1..125506
/organism="Felis catus"
/db_xref="taxon:9685"
/clone="RP86-400G2"
/clone_lib="RP86"
1..27057
/notes="assembly_fragment"
clone_end:SP6
vector_side:left
27158..109603
/notes="assembly_fragment"
109704..118983
/notes="assembly_fragment"
119084..125506
/notes="assembly_fragment"
clone_end:T7
vector_side:right

```

##### misc\_feature

```

1..27057
/notes="assembly_fragment"
clone_end:SP6
vector_side:left
27158..109603
/notes="assembly_fragment"
109704..118983
/notes="assembly_fragment"
119084..125506
/notes="assembly_fragment"
clone_end:T7
vector_side:right

```

##### misc\_feature

```

109704..118983
/notes="assembly_fragment"
119084..125506
/notes="assembly_fragment"

```

##### misc\_feature

```

119084..125506
/notes="assembly_fragment"
clone_end:T7
vector_side:right

```

```

BASE COUNT 31320 a 33382 c 32878 g 27625 t 301 others
ORIGIN

```

```

Query Match 74.8%; Score 17.2; DB 2; Length 125506;
Best Local Similarity 86.4%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 ACCAGGCGTCTCGTGGCCACA 22
||||||| ||||| ||||| ||||| |||||

```

```

Db 50292 ACCAGGCGTCTCGAGTCCACA 50271

```

#### RESULT 19

##### AC100788/c

```

LOCUS AC100788 127568 bp DNA linear HTG 21-AUG-2002
DEFINITION Homo sapiens chromosome 17 clone CTD-2373H9 map 17, *** SEQUENCING
IN PROGRESS ***, 4 unordered pieces.

```

##### ACCESSION

```

AC100788
AC100788.2 GI:22381955

```

##### KEYWORDS

```

HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.

```

##### SOURCE

##### ORGANISM

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

##### REFERENCE

```

1 (bases 1 to 127568)
Birren,B., Nusbaum,C. and Lander,E.

```

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 127568)

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A., Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gird, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Aug 21, 2002 this sequence version replaced gi:17048158.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L20504

Center clone name: 2373\_H\_9

\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 49542: contig of 49542 bp in length  
\* 49543 49642: gap of 100 bp  
\* 49643 67854: contig of 18212 bp in length  
\* 67855 67954: gap of 100 bp  
\* 67955 74191: contig of 6237 bp in length  
\* 74192 74291: gap of 100 bp  
\* 74292 127568: contig of 53277 bp in length.

FEATURES  
source

1..127568  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="CTD-2373H9"  
/clone\_lib="CITD1 Human BAC"  
29694 a 34781 c 33912 g 28702 t 479 others

Query Match 74.8%; Score 17.2; DB 2; Length 127568;  
Best Local Similarity 82.6%; Pred. No. 6.4e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACCAGGCGTCTCGTGGCCACAT 23  
||||| ||||||| |||||||  
Db 124475 ACCAGGCGTCTCGTGGCCACAT 124453

RESULT 20  
AC017124  
LOCUS 143450 bp DNA linear HTG 09-DEC-1999  
DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
ACCESSION AC017124  
VERSION AC017124.1 GI:6553862  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS Adams, M. and Venter, J.C.  
TITLE Direct Submission  
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT This sequence was identified as CDM:10209560 by the submitter. For more information on this record e-mail to [fly@celera.com](mailto:fly@celera.com).

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

FEATURES

source

1..143450

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

BASE COUNT 41368 a 30430 c 30627 g 41025 t

ORIGIN

Query Match 74.8%; Score 17.2; DB 2; Length 143450;  
Best Local Similarity 86.4%; Pred. No. 6.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCAGGCGTCTCGTGGCCACAT 23

||| ||||||| |||||||

Db 112908 CCGGACGTCTCGTGGCCACAT 112929

RESULT 21

HS1164K7/c

LOCUS 154076 bp DNA linear HTG 22-OCT-2001

DEFINITION Homo sapiens chromosome 3 clone BAC RPCI-11 1164K7 map 3q21, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 2 ordered pieces.

ACCESSION AL592209

VERSION AL592209.2 GI:16415917

KEYWORDS HTG; HTGS\_PHASE2; HTGS\_ACTIVEFIN; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 154076)

REFERENCE

AUTHORS Sudbrak, R., Borzym, K., Koslura, A., Rak, K., Omran, H., Lehrach, H. and Reinhardt, R.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 154076)

AUTHORS MOLGENR.

TITLE Direct Submission

JOURNAL Submitted (21-JUN-2001) MPIMG, Abt. Lehrach, Max Planck Institut

Fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195 Germany

On Oct 24, 2001 this sequence version replaced gi:14530799.

contig 01 1..149871

contig 02 149972..154076

Clone received from Resource Centre of the Human Genome Project at

the Max-Planck-Institut for Molecular Genetics.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 2 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 149871: contig of 149871 bp in length  
\* 149872 149971: gap of 100 bp  
\* 149972 154076: contig of 4105 bp in length.

FEATURES  
source

Location/Qualifiers  
1..154076  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/map="3q21"  
/clone="BAC RPCI-11 1164K7"  
/clone\_lib="RPCI-11 Human BAC library, originating  
institute: Roswell Park Cancer Institute"  
/note="region between markers D3S3607-D3S1290"  
BASE COUNT 46174 a 29448 c 30349 g 48005 t 100 others  
ORIGIN

Query Match 74.8%; Score 17.2; DB 2; Length 154076;  
Best Local Similarity 86.4%; Pred. No. 6.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CCAGCGTCTCGTGGCCACAT 23

Db 112830 CCAGCGTCTCATGAGCCACAT 112809

RESULT 22

AC117409  
LOCUS AC117409 154336 bp DNA linear PRI 14-AUG-2002  
DEFINITION Homo sapiens 3 BAC RP11-64K7 (Roswell Park Cancer Institute Human  
BAC Library) complete sequence.  
AC117409  
VERSION AC117409.4 GI:22218445  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 154336)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
Albrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carroll,T.P., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
Loulseghe,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,  
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,  
Neelson,D., Newtonson,J., Newtonson,N., Nguyen,N., Nguyen,N.,

Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,  
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,  
Pickens,R., Primus,E., Pu,L.D., Quiles,M., Ren,Y., Rives,M.,  
Rojas,A., Rojubokan,I., Rolfe,S., Ruiz,S., Savery,G., Scherer,S.,  
Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,  
Sonake,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,  
Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,  
Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,  
Vara,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,  
Warren,R., Washington,C., Watlington,S., Williams,G.,  
Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,  
Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and  
Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 154336)

Worley,K.C.

Direct Submission

Submitted (10-APR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 154336)

Worley,K.C.

Direct Submission

Submitted (04-JUN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 154336)

Worley,K.C.

Direct Submission

Submitted (14-AUG-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Aug 14, 2002 this sequence version replaced gi:21314809.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE:

Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:

This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annot>  
ation.html.

FEATURES

Location/Qualifiers

1..154336

/organism="Homo sapiens"





TITLE	Direct Submission	Human Genome Sequencing Center, Department
JOURNAL	Submitted (22-SEP-1999)	of Molecular and Human Genetics, Baylor College of Medicine, One
REFERENCE	5	Baylor Plaza, Houston, TX 77030, USA
AUTHORS	(bases 1 to 178242)	
TITLE	Worley, K.C.	
JOURNAL	Direct Submission	Human Genome Sequencing Center, Department
	Submitted (09-OCT-1999)	of Molecular and Human Genetics, Baylor College of Medicine, One
REFERENCE	6	Baylor Plaza, Houston, TX 77030, USA
AUTHORS	(bases 1 to 178242)	
TITLE	Worley, K.C.	
JOURNAL	Direct Submission	Human Genome Sequencing Center, Department
	Submitted (13-JUN-2000)	of Molecular and Human Genetics, Baylor College of Medicine, One
REFERENCE	7	Baylor Plaza, Houston, TX 77030, USA
AUTHORS	(bases 1 to 178242)	
TITLE	Worley, K.C.	
JOURNAL	Direct Submission	Human Genome Sequencing Center, Department
	Submitted (28-FEB-2001)	of Molecular and Human Genetics, Baylor College of Medicine, One
REFERENCE	91	Baylor Plaza, Houston, TX 77030, USA
AUTHORS	On or before Jun 13, 2000	this sequence version replaced
TITLE	gi:5668751, gi:5668753.	
JOURNAL	INFORMATION: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a>	or email
COMMENT	qc-help@bcm.tmc.edu	

**CLONE LENGTH:** This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

**ANNOTATION OF FEATURES:**  
 STSSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.  
 Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished,) for Human and Mouse sequences.  
 Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < le-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

```

QUALSTAT-REPORT-----
----- Summary Statistics -----
Contig length:                178243
Phrap values in estimate:     177636
Average error rate (BCW-Phrap estimate): 3.24893e-05
Fraction of Phrap values less than 40 : 0.00339458
Number of consensus changing edits: 10
Number of N's in consensus : 0

```

	Position	Original-Context	Consensus Changing edits	Edited-Context
	43033	ttgcccctgg(n)ccccgcggc		ttgccctcgg(c)ccccgcggc
	60652	cagcctggcc(n)actngtgaa		cagcctggcc(a)acaagtgaac
	60655	cctggccnac(t)nggtgaacc		cctggccaac(a)agtgaaacc
	60656	ttagccnact(n)ggtgaacc		ctggccaaca(a)ggtgaacc
	60707	ttgcccggg(n)ngtggcta		ctggcccggg(c)gtggcgcta
	60709	agcccggg(n)ggtggcta		agcccggcg(c)ggtggcta
	60819	attatatagg(n)gggacctatg		attatatagg(t)gggacctatg
	106499	gagctatggg(n)aaaagacagt		gagctatggg(g)aaaagacagt
	118211	cagtgcgcc(t)aaaactctg		cagtgcgcc(a)aaaactctg
	133724	ttttttttt(n)gatacagggt		ttttttttt(t)gatacagggt

  

----- Distribution of Quality < 40 Bases -----

# bases	5	10	15	20	25	30	35	40
	5001							
	4501							
	4001							
	3501							
	3001							
	2501							
	2001					*		
	1501					*		
	1001					*		
	501		*	*	*	*	*	*
	01	*	*	*	*	*	*	*

Phrap Value Range

  

Version: 1.01 xfqo.

FEATURES	Location/Qualifiers
source	1..178242
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="3q25-26"
	/clone="CITB-243A6"
repeat_region	complement(155..289)
	/rpt_family="MER5B"
repeat_region	complement(378..413)
	/rpt_family="MER102"
repeat_region	complement(455..765)
	/rpt_family="AluX"
repeat_region	complement(956..1064)
	/rpt_family="MIR"
repeat_region	1531..1722
	/rpt_family="MLT1D"
repeat_region	complement(1748..1963)
	/rpt_family="MLT1C"
repeat_region	1964..2014
	/rpt_family="(TG)n"
repeat_region	complement(2015..2230)
	/rpt_family="MLT1C"
repeat_region	2577..2712
	/rpt_family="L2"
repeat_region	3768..3916
	/rpt_family="MER58A"
repeat_region	complement(3925..3998)
	/rpt_family="MER45B"
repeat_region	4807..4844
	/rpt_family="MIR"
repeat_region	complement(5947..6261)
	/rpt_family="MLT1F"
repeat_region	complement(6301..6343)
	/rpt_family="MLT1F"
repeat_region	7176..7408
	/rpt_family="L2"
repeat_region	7486..7735
	/rpt_family="L2"
repeat_region	complement(8255..8359)



```

/rpt_family="MIR"
Join(8735..8895,20435..21674)
/gene="Homo sapiens caveolin-3 (CAV3) mRNA, complete cds,
AF043101 and AF036365"
gene
Join(8741..8895,52375..52434,116209..116631)
/gene="Unigene cluster containing AA448334 and AA452471"
repeat_region
complement(9628..9917)
/rpt_family="Aluub"

```

```

Query Match          74.8%; Score 17.2; DB 9; Length 178242;
Best Local Similarity 86.4%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACCAGCGTCTCGTGGCCACA 22
||||| ||||| ||||| ||||| |||||
Db 123607 ACCATGGGTACGTGGCCACA 123628

```

```

RESULT 24
AC099027 185425 bp DNA linear INV 08-NOV-2001
LOCUS Drosophila melanogaster, chromosome 2R, region 53C-53D, BAC clone
DEFINITION BACR06115, complete sequence.
ACCESSION AC099027
VERSION AC099027.1 GI:16798959
KEYWORDS HTG.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster

```

```

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
Celinker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Fierliera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Schaefer,F.,
Shouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 2R, region 53C-53D
Unpublished
2 (bases 1 to 185425)

```

```

REFERENCE
AUTHORS Celinker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Fierliera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Schaefer,F.,
Shouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 2R, region 53C-53D
Unpublished
2 (bases 1 to 185425)

```

```

TITLE JOURNAL
REFERENCE
AUTHORS Celinker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Fierliera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Shouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirska,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Direct Submission
Submitted (08-NOV-2001) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

```

```

TITLE JOURNAL
COMMENT
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence

```

```

archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu.
Location/Qualifiers
1..185425
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="53C-53D"
/clone="BACR06115 (D1338)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
pBACE3.6)"
BASE COUNT 53519 a 39068 c 39245 g 53593 t
ORIGIN

```

```

Query Match          74.8%; Score 17.2; DB 3; Length 185425;
Best Local Similarity 86.4%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCAGCGTCTCGTGGCCACAT 23
||||| ||||| ||||| ||||| |||||
Db 145002 CCGGAGCTCTCGTGGCCACAT 145023

```

```

RESULT 25
AC122284 194991 bp DNA linear HTG 23-MAY-2002
LOCUS Mus musculus chromosome UNK clone RP23-246A5, WORKING DRAFT
DEFINITION AC122284
ACCESSION AC122284.1 GI:21105140
VERSION AC122284
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus.
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS 1 (bases 1 to 194991)
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
2 (bases 1 to 194991)
AUTHORS McPherson,J.D. and Waterston,R.H.
DIRECT SUBMISSION
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT

```

```

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BA0246A05
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 190912 bases at least Q40
Consensus quality: 192510 bases at least Q30
Consensus quality: 193557 bases at least Q20
Insert size: 204000; agarose-fp
Quality coverage: 11.70 in Q20 bases; agarose-fp
Quality coverage: 6.63 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

```

\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1173: contig of 1173 bp in length  
1174 1273: gap of unknown length  
1274 2330: contig of 1057 bp in length  
2331 2430: gap of unknown length  
2431 3593: contig of 1163 bp in length  
3594 4999: gap of unknown length  
4999 5099: contig of 1306 bp in length  
5099 6465: contig of 1366 bp in length  
6466 6565: gap of unknown length  
6566 9541: contig of 2976 bp in length  
9542 9641: gap of unknown length  
9642 13055: contig of 3414 bp in length  
13056 13155: gap of unknown length  
13156 20585: contig of 7430 bp in length  
20586 20685: gap of unknown length  
20686 29723: contig of 9038 bp in length  
29724 29823: gap of unknown length  
29824 43883: contig of 14060 bp in length  
43884 43983: gap of unknown length  
43984 56074: contig of 12091 bp in length  
56075 56174: gap of unknown length  
56175 80163: contig of 23989 bp in length  
80164 80263: gap of unknown length  
80264 135442: contig of 55079 bp in length  
135443 135443: gap of unknown length  
135443 194991: contig of 59549 bp in length.

Location/Qualifiers

1. 194991  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="UNK"  
/clone="RP23-246A5"

misc\_feature

1. 1173  
/note="assembly\_name:Contig9"  
1274..2330  
/note="assembly\_name:Contig47"  
2431..3593  
/note="assembly\_name:Contig50"  
3694..4999  
/note="assembly\_name:Contig53"  
5100..6465  
/note="assembly\_name:Contig57"  
6566..9541  
/note="assembly\_name:Contig58"  
9642..13055  
/note="assembly\_name:Contig59  
clone\_end:SP6  
vector\_side:left"  
13156..20585  
/note="assembly\_name:Contig60"  
20686..29723  
/note="assembly\_name:Contig61"  
29824..43883  
/note="assembly\_name:Contig62"  
43984..56074  
/note="assembly\_name:Contig63"  
56175..80163  
/note="assembly\_name:Contig64  
clone\_end:T7  
vector\_side:left"  
80264..135442  
/note="assembly\_name:Contig65"  
135443..194991  
/note="assembly\_name:Contig66"

misc\_feature

13156..20585  
/note="assembly\_name:Contig60"  
20686..29723  
/note="assembly\_name:Contig61"  
29824..43883  
/note="assembly\_name:Contig62"  
43984..56074  
/note="assembly\_name:Contig63"  
56175..80163  
/note="assembly\_name:Contig64  
clone\_end:T7  
vector\_side:left"  
80264..135442  
/note="assembly\_name:Contig65"  
135443..194991  
/note="assembly\_name:Contig66"

misc\_feature

13156..20585  
/note="assembly\_name:Contig60"  
20686..29723  
/note="assembly\_name:Contig61"  
29824..43883  
/note="assembly\_name:Contig62"  
43984..56074  
/note="assembly\_name:Contig63"  
56175..80163  
/note="assembly\_name:Contig64  
clone\_end:T7  
vector\_side:left"  
80264..135442  
/note="assembly\_name:Contig65"  
135443..194991  
/note="assembly\_name:Contig66"

misc\_feature

13156..20585  
/note="assembly\_name:Contig60"  
20686..29723  
/note="assembly\_name:Contig61"  
29824..43883  
/note="assembly\_name:Contig62"  
43984..56074  
/note="assembly\_name:Contig63"  
56175..80163  
/note="assembly\_name:Contig64  
clone\_end:T7  
vector\_side:left"  
80264..135442  
/note="assembly\_name:Contig65"  
135443..194991  
/note="assembly\_name:Contig66"

BASE COUNT 55728 a 42930 c 41683 g 53332 t 1318 others  
ORIGIN

Query Match 74.8% Score 17.2; DB 2; Length 194991;

Best Local Similarity 86.4%; Pred. No. 6.5e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ACCAGGCGTCTCGTGGGCCACA 22

Db 21192 ACCAGAGGTCGTGGGCCACA 21213

Search completed: December 2, 2002, 22:01:00  
Job time : 3434 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2002, 20:55:23 ; Search time 260 Seconds  
(without alignments)  
199.215 Million cell updates/sec

Title: US-09-700-906a-3

Perfect score: 23

Sequence: 1 accagggtctctgtggccacat 23

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 2059836

Minimum DB seq length: 66

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	23	100.0	11435	21	AAA64659
C 2	23	100.0	12493	21	AAZ43872
C 3	23	100.0	12515	21	AAA64658
C 4	23	100.0	12515	24	ABL69122
5	17.2	74.8	189	22	ABA75966
6	17.2	74.8	189	22	ABA40530
7	17.2	74.8	189	22	AAK24644
8	17.2	74.8	189	22	AAK50641
9	17.2	74.8	189	22	AAI27655

10	17.2	74.8	189	22	AAI56622	Probe #25308 used
11	17.2	74.8	189	24	ABA24117	Human genome-deriv
12	17.2	74.8	573	22	ABA63634	Human foetal liver
13	17.2	74.8	573	22	ABA30830	Probe #9296 for ge
14	17.2	74.8	573	22	AAK12157	Human brain expres
15	17.2	74.8	573	22	AAK37875	Human bone marrow
16	17.2	74.8	573	22	AAI18637	Probe #8570 for ge
17	17.2	74.8	573	22	AAI43749	Probe #12435 used
18	17.2	74.8	573	24	ABS11872	Human genome-deriv
19	17.2	74.8	1143	21	AAZ65257	Human secreted pro
20	17.2	74.8	1851	23	ABL07567	Drosophila melanog
21	17.2	74.8	7832	23	ABL07566	Bacillus lichenifo
22	16.8	73.0	1557	24	ABK72958	Human DNase I gene
23	16.8	73.0	4042	17	AAI35522	Human reproductive
24	16.8	73.0	25701	22	AAI07078	Human secreted exp
25	16.8	73.0	25758	22	AAI07077	Human secreted exp
26	16.6	72.2	354	21	AAA44913	Human secreted pro
27	16.6	72.2	363	20	AAI41003	Human secreted pro
28	16.6	72.2	373	20	AAI51631	Human secreted pro
29	16.6	72.2	391	21	AAI16890	Human secreted pro
30	16.6	72.2	451	20	AAI41156	Human secreted pro
31	16.6	72.2	716	22	AAI94017	Primer specific fo
32	16.6	72.2	1113	22	AAI61225	Borna disease viru
33	16.6	72.2	1755	22	AAI93818	Human cDNA encodin
34	16.6	72.2	2000	22	AAI51821	Human polynucleoti
35	16.6	72.2	2029	21	AAI298172	Human signal pepti
36	16.6	72.2	2313	19	AAI22687	DNA encoding a CD4
37	16.6	72.2	2324	22	ABA09223	Human PRO263 homol
38	16.6	72.2	2369	24	ABL90698	Human polynucleoti
39	16.6	72.2	2372	20	AAI52250	Protein PRO263 CDN
40	16.6	72.2	2372	22	AAI92060	Human PRO263 cDNA.
41	16.6	72.2	2372	22	AAI72408	Human PRO263 cDNA.
42	16.6	72.2	2372	24	ABL95576	Human angiogenesis
43	16.6	72.2	2372	24	ABL88087	Human PRO263 cDNA
44	16.6	72.2	2967	21	AAI59803	Human secreted pro
45	16.6	72.2	3517	18	AAI61291	Mouse truncated be

ALIGNMENTS

RESULT 1  
AAA64659/c  
ID AAA64659 standard; cDNA; 11435 BP.

AC AAA64659;

DT 02-JAN-2001 (first entry)

XX cDNA sequence encoding the antigen of monoclonal antibody Ki-67.

DE Melanoma; thyroid tumour; MEL3; Thy5; Thy6; Thy11; Thy14; Thy15;  
KW rectal cancer; lung cancer; breast cancer; colon cancer;  
KW antibody Ki-67; ss.

XX Homo sapiens.

OS WO200050595-A2.

PN 31-AUG-2000.

PD 25-FEB-2000; 2000WO-US04929.

PF 25-FEB-1999; 99US-0257417.

PR (GOUT/) GOUT I.

PA (ROD/) RODIN N.

PA (FILO/) FILOENKO V.

PA (MATS/) MATSUKA G.

PA (SCAN/) SCANLAN M.

PA (OLDL/) OLD L.

PA (BILY/) BILYNSKY B.

PI Gout I, Rodnin N, Filonenko V, Matsuka G, Scanlan M, Old L;  
PI Bilynsky B;  
XX  
DR WPI: 2000-572092/53.  
XX  
PT Novel isolated nucleic acid molecules for diagnosing and treating  
PT melanoma, thyroid tumors, rectal, lung, breast and colon cancers -  
XX  
XX  
PS Claim 17; Page 81-85; 94pp; English.  
XX  
CC The specification describes polynucleotides which are associated with  
CC melanoma and thyroid tumors. Specifically, these are MEL3, MEL7, Thy5,  
CC Thy6, Thy11, Thy14, and Thy15. The polynucleotides are useful for  
CC diagnosing and treating a patient with melanoma, thyroid tumour,  
CC rectal cancer, lung cancer, breast cancer or colon cancer. The  
CC present sequence represents a polynucleotide of the invention.  
XX  
XX  
SQ Sequence 11435 BP; 3802 A; 2828 C; 2684 G; 2121 T; 0 other;  
  
Query Match 100.0%; Score 23; DB 21; Length 11435;  
Best Local Similarity 100.0%; Pred. No. 0.19;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 ACCAGGCGTCTCGTGGCCACAT 23  
Db 219 ACCAGGCGTCTCGTGGCCACAT 197  
  
RESULT 2  
AAZ43872/c  
ID AA243872 standard; DNA; 12493 BP.  
XX  
AC AA243872;  
XX  
DT 10-MAR-2000 (first entry)  
XX  
DE Human cell cycle protein Ki-67 DNA.  
XX  
XX  
KW Cell cycle protein; Ki-67; therapy; cell proliferation; allergy;  
KW tumor treatment; autoimmune disease; scar formation; inflammation;  
KW rheumatic disease; transplantation; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 197..9567  
FT /\*tag= a  
FT /product= "Ki-67"  
XX  
PN DE19822954-A1.  
XX  
XX 25-NOV-1999.  
XX  
XX 22-MAY-1998; 98DE-1022954.  
XX  
XX 22-MAY-1998; 98DE-1022954.  
XX  
XX (BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.  
XX  
PI Flad H, Gerdes J, Boehle A, Deinert I;  
XX  
XX WPI: 2000-039964/04.  
DR P-PSDB; AAY50976.  
XX  
XX Ki-67 gene antisense oligonucleotide -  
PT  
XX  
PS Disclosure; Page 5-19; 36pp; German.  
XX  
XX This invention describes a novel oligoribo- or oligodeoxyribonucleotide,  
CC characterized in that, it hybridizes to mRNA that encodes protein Ki-67  
CC at a physiologically acceptable salt concentration. The oligoribo- or  
CC oligodeoxyribonucleotide which is complementary to Ki-67, a protein  
CC active at all stages of the cell cycle except G<sub>0</sub>, is useful for therapy

CC of illnesses with increased cell proliferation and particularly for  
CC treatment of tumors, autoimmune diseases, scar formation, inflammation,  
CC allergy, rheumatic diseases and defence against transplantation. This  
CC sequence encodes the human cell cycle protein Ki-67 which is described in  
CC the method of the invention.  
XX  
XX  
SQ Sequence 12493 BP; 4143 A; 3048 C; 2929 G; 2373 T; 0 other;  
  
Query Match 100.0%; Score 23; DB 21; Length 12493;  
Best Local Similarity 100.0%; Pred. No. 0.19;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 ACCAGGCGTCTCGTGGCCACAT 23  
Db 219 ACCAGGCGTCTCGTGGCCACAT 197  
  
RESULT 3  
AAA64658/c  
ID AAA64658 standard; cDNA; 12515 BP.  
XX  
AC AAA64658;  
XX  
DT 02-JAN-2001 (first entry)  
XX  
DE cDNA sequence encoding the antigen of monoclonal antibody Ki-67.  
XX  
KW Melanoma; thyroid tumour; MEL3; MEL7; Thy5; Thy6; Thy11; Thy14; Thy15;  
KW rectal cancer; lung cancer; breast cancer; colon cancer;  
KW antibody Ki-67; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO2000050595-A2.  
XX  
PD 31-AUG-2000.  
XX  
XX 25-FEB-2000; 2000WO-US04929.  
XX  
XX 25-FEB-1999; 99US-0257417.  
XX  
XX (GOUT/) GOUT I.  
PA (RODN/) RODNIN N.  
PA (FILO/) FILONENKO V.  
PA (MATS/) MATSUKA G.  
PA (SCAN/) SCANLAN M.  
PA (OLDL/) OLD L.  
PA (BILY/) BILYNSKY B.  
XX  
XX Gout I, Rodnin N, Filonenko V, Matsuka G, Scanlan M, Old L;  
PI Bilynsky B;  
XX  
XX WPI: 2000-572092/53.  
XX  
XX Novel isolated nucleic acid molecules for diagnosing and treating  
PT melanoma, thyroid tumors, rectal, lung, breast and colon cancers -  
XX  
XX Claim 17; Page 77-81; 94pp; English.  
XX  
CC The specification describes polynucleotides which are associated with  
CC melanoma and thyroid tumors. Specifically, these are MEL3, MEL7, Thy5,  
CC Thy6, Thy11, Thy14, and Thy15. The polynucleotides are useful for  
CC diagnosing and treating a patient with melanoma, thyroid tumour,  
CC rectal cancer, lung cancer, breast cancer or colon cancer. The  
CC present sequence represents a polynucleotide of the invention.  
XX  
XX  
SQ Sequence 12515 BP; 4164 A; 3050 C; 2928 G; 2373 T; 0 other;  
  
Query Match 100.0%; Score 23; DB 21; Length 12515;  
Best Local Similarity 100.0%; Pred. No. 0.19;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 ACCAGGCGTCTCGTGGCCACAT 23

```
Db 219 ACCAGGCGTCTCGTGGCCACAT 197
|||||
RESULT 4
ABL69122/c
ID ABL69122 standard; DNA; 12515 BP.
XX
AC ABL69122;
XX
XX 15-MAY-2002 (first entry)
XX
DE Kidney cancer related gene sequence SEQ ID NO:7459.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
XX WO200194629-A2.
XX
XX 13-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233133P.
PR 20-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
PA Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX
PI Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 7459; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic
XX activity and can be used in gene therapy. M1 can be used for screening
XX an anti-neoplastic agent, and can be used for producing a product which
XX is the data collected with respect to the anti-neoplastic agent as a
XX result of M1, and the data is sufficient to convey the chemical
XX structure and/or properties of the agent. M1 can be used in the
XX treatment of cancer such as colon, breast, stomach, lung, thyroid,
XX oesophageal, ovarian, kidney, prostate or pancreatic cancer,
XX adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
XX infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
XX carcinoma, papillary carcinoma and Wilm's tumour.
XX
XX Sequence 12515 BP; 4166 A; 3048 C; 2928 G; 2373 T; 0 other;
SQ
Query Match 100.0%; Score 23; DB 24; Length 12515;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACCAGGCGTCTCGTGGCCACAT 23
Db 219 ACCAGGCGTCTCGTGGCCACAT 197
|||||
RESULT 5
ABA75966
ID ABA75966 standard; DNA; 189 BP.
XX
XX ABA75966;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #24271.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
```

PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human fetal liver -  
PS Claim 4; SEQ ID NO 24271; 639pp + sequence listing; English.  
XX  
CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 189 BP; 52 A; 57 C; 40 G; 40 T; 0 other;  
Query Match 74.8%; Score 17.2; DB 22; Length 189;  
Best Local Similarity 86.4%; Pred. No. 84;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ACCAGGGCTCTCGTGGGCCACA 22  
||||| ||||| ||||| ||||| |||||  
Db 1 ACCATGGGTACGTTGGGCCACA 22  
RESULT 6  
ABA40530  
ID ABA40530 standard; DNA; 189 BP.  
XX  
AC ABA40530;  
XX  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Probe #18996 for gene expression analysis in human heart cell sample.  
XX  
KW Human; gene expression; heart; microarray; vascular system; probe;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00666.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488899/53.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
PS Claim 4; SEQ ID No 18996; 530pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes for  
XX measuring human gene expression in a sample derived from human heart. The  
XX present sequence is one such probe. The probes may be used for  
XX predicting, measuring and displaying gene expression in samples derived  
XX from the human heart via microarrays. By measuring gene expression, the  
XX probes are useful for predicting, diagnosing, grading, staging,

CC monitoring and prognosing diseases of the human heart and vascular system  
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
CC congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 189 BP; 52 A; 57 C; 40 G; 40 T; 0 other;  
Query Match 74.8%; Score 17.2; DB 22; Length 189;  
Best Local Similarity 86.4%; Pred. No. 84;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ACCAGGGCTCTCGTGGGCCACA 22  
||||| ||||| ||||| ||||| |||||  
Db 1 ACCATGGGTACGTTGGGCCACA 22  
RESULT 7  
AAK24644  
ID AAK24644 standard; DNA; 189 BP.  
XX  
AC AAK24644;  
XX  
XX 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe SEQ ID NO: 24635.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00667.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
PS Example 4; SEQ ID NO: 24635; 650pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX epilepsy and cancers. The present sequence is one of the probes of the  
XX invention.  
XX  
SQ Sequence 189 BP; 52 A; 57 C; 40 G; 40 T; 0 other;  
Query Match 74.8%; Score 17.2; DB 22; Length 189;  
Best Local Similarity 86.4%; Pred. No. 84;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACCAGCGTCTCGTGGGCACA 22  
 ||||| | ||| ||||| |||||  
 Db 1 ACCATGGGTCACGTGGGCACA 22

## RESULT 8

AAK50641  
 ID AAK50641 standard; DNA; 189 BP.

XX AC  
 XX AAK50641;

DT 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 25198.

DE Human bone marrow expressed exon; gene expression analysis; probe;  
 KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX OS Homo sapiens.

XX WO200157276-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000668.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human bone marrow -

XX Example 4; SEQ ID NO: 25198; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers

CC such as lymphoma, leukaemia and myeloma. The present sequence is one of

XX the probes of the invention.

XX SQ Sequence 189 BP; 52 A; 57 C; 40 G; 40 T; 0 other;

## RESULT 9

AAI27655  
 ID AAI27655 standard; DNA; 189 BP.

XX AC  
 XX AAI27655;

DT 12-OCT-2001 (first entry)

XX Probe #17588 for gene expression analysis in human cervical cell sample.

XX

KW Probe; human; microarray; gene expression; cervical epithelial cell;  
 KW cervical cancer; ss.

OS Homo sapiens.

XX WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000670.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human cervical epithelial cells -

XX Claim 25; SEQ ID No 17588; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes  
 CC (SENPs). The present sequence is one such probe. The SENPs are derived  
 CC from human HeLa cells. The SENPs can be used to produce a single exon  
 CC microarray, which can be used for measuring human gene expression in a  
 CC sample derived from human cervical epithelial cells. By measuring gene  
 CC expression, the probes are therefore useful in grading and/or staging  
 CC of diseases of the cervix, notably cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 189 BP; 52 A; 57 C; 40 G; 40 T; 0 other;

Query Match 74.8%; Score 17.2; DB 22; Length 189;  
 Best Local Similarity 86.4%; Pred. No. 84;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACCAGCGTCTCGTGGGCACA 22  
 ||||| | ||| ||||| |||||

Db 1 ACCATGGGTCACGTGGGCACA 22

## RESULT 10

AAI56622  
 ID AAI56622 standard; DNA; 189 BP.

XX AC  
 XX AAI56622;

DT 17-OCT-2001 (first entry)

XX Probe #25308 used to measure gene expression in human placenta sample.  
 DE Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder; ss.

XX OS Homo sapiens.

XX WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000663.

XX

```
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta.
XX
XX Claim 25; SEQ ID No 25308; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 189 BP; 52 A; 57 C; 40 G; 40 T; 0 other;
XX
XX Query Match 74.8%; Score 17.2; DB 22; Length 189;
XX Best Local Similarity 86.4%; Pred. No. 84;
XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
Qy 1 ACCAGGCGTCTCGTGGGCCACA 22
Db 1 ACCATGGGTACGTGGGCCACA 22
XX
RESULT 11
AB2524117
ID ABS24117 standard; DNA; 189 BP.
AC ABS24117;
XX
XX 19-AUG-2002 (first entry)
XX
XX Human genome-derived single exon probe ORF from lung SEQ ID No 24108.
XX
XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenier syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease; open reading frame; ORF.
XX
XX Homo sapiens.
XX
XX OS
XX WO200186003-A2.
XX
XX 15-NOV-2001.
XX
XX 30-JAN-2001; 2001WO-US00665.
XX
XX 04-FEB-2000; 2000US-180312P.
XX 26-MAY-2000; 2000US-207456P.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-234687P.
XX 27-SEP-2000; 2000US-236359P.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
```

```
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples.
XX
XX Claim 4; SEQ ID No 24108; 634pp; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human lung comprising single exon nucleic acid probes having one of
XX 12614 nucleic acid sequences mentioned in the specification, or their
XX complements or the 12387 open reading frames derived from the 12614
XX probes. Also included are a microarray comprising the novel set of
XX probes; the novel set of probes which hybridise at high stringency to a
XX nucleic acid expressed in the human lung; measuring gene expression in a
XX sample derived from human lung, comprising (a) contacting the array with
XX a collection of detectably labeled nucleic acids derived from human lung
XX mRNA, and (b) measuring the label detectably bound to each probe of
XX the array; identifying exons in a eukaryotic genome, comprising
XX (a) algorithmically predicting at least one exon from genomic sequences
XX of the eukaryote; and (b) detecting specific hybridisation of detectably
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
XX having a fragment identical to the predicted exon, the probe is included
XX in the above mentioned microarray; assigning exons to a single gene,
XX comprising (a) identifying exons from genomic sequence by the method
XX above and (b) measuring the expression of each of the exons in several
XX tissues and/or cell types using hybridisation to a single exon
XX microarrays having a probe with the exon, where a common pattern of
XX expression of the exons in the tissues and/or cell types indicates that
XX the exons should be assigned to a single gene; a peptide comprising one
XX of 12011 sequences, mentioned in the specification, or encoded by the
XX probes/open reading frames (ORF). The probes are used for gene
XX expression analysis, and for identifying exons in a gene, particularly
XX using human lung derived mRNA and for the study of lung diseases
XX such as asthma, lung cancer, chronic obstructive pulmonary disease
XX (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
XX fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
XX Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
XX haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
XX pulmonary alveolar proteinosis, Karagenier syndrome, fibrocystic
XX pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
XX and hyaline membrane disease. The present sequence is a single exon
XX probe open reading frame of the invention.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 189 BP; 52 A; 57 C; 40 G; 40 T; 0 other;
XX
XX Query Match 74.8%; Score 17.2; DB 24; Length 189;
XX Best Local Similarity 86.4%; Pred. No. 84;
XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
Qy 1 ACCAGGCGTCTCGTGGGCCACA 22
Db 1 ACCATGGGTACGTGGGCCACA 22
XX
RESULT 12
ABA63634
ID ABA63634 standard; DNA; 573 BP.
XX
XX ABA63634;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #11939.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
```







```

RESULT 17
AAI43749
ID AAI43749 standard; DNA; 573 BP.
XX
AC AAI43749;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #12435 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI: 2001-488897/53.
XX
Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 12435; 654pp; English.
XX
The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 573 BP; 147 A; 162 C; 116 G; 148 T; 0 other;

Query Match 74.8%; Score 17.2; DB 22; Length 573;
Best Local Similarity 86.4%; Pred. No. 90;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACCAGGCGTCTGTTGGGCACCA 22
||||| | ||| ||||| |||||
Db 267 ACCATGGGTCACGTGGGCACA 288

RESULT 18
ABS11872
ID ABS11872 standard; DNA; 573 BP.
XX
AC ABS11872;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe from lung SEQ ID No 11863.
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

```

```

KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI: 2002-114183/15.
XX
Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
PS Claim 1; SEQ ID No 11863; 634pp; English.
XX
The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a single exon
CC probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

```

XX SQ Sequence 573 BP; 147 A; 162 C; 116 G; 148 T; 0 other;

Query Match 74.8%; Score 17.2; DB 24; Length 573;  
 Best Local Similarity 86.4%; Pred. No. 90;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ACCAGCGTCTCGTGGGCACCA 22  
 ||||| | ||| ||||| |||||  
 Db 267 ACCATGGTCTACGTGGGCACA 288

RESULT 19  
 AAZ65257  
 ID AAZ65257 standard; DNA; 1143 BP.

XX AC AAZ65257;  
 XX AC AAZ65257;  
 DT 23-MAR-2000 (first entry)  
 XX Human secreted protein gene 8.

XX Human; secreted protein; cancer; tumour; developmental abnormality;  
 KW foetal deficiency; blood disorder; immune system disorder; inflammation;  
 KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;  
 KW schizophrenia; arthritis; psoriasis; sepsis; skin disorder;  
 KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;  
 KW digestive disorder; endocrine disorder; infection; AIDS; leukaemia;  
 KW therapy; chromosome 10; ds.

XX Homo sapiens.  
 OS Homo sapiens.  
 XX WO9958660-A1.  
 PN 18-NOV-1999.

XX PD 06-MAY-1999; 99WO-US09847.  
 XX 12-MAY-1998; 98US-0085093.  
 PR 12-MAY-1998; 98US-0085094.  
 PR 12-MAY-1998; 98US-0085105.  
 PR 12-MAY-1998; 98US-0085180.  
 PR 18-MAY-1998; 98US-0085906.  
 PR 18-MAY-1998; 98US-0085920.  
 PR 18-MAY-1998; 98US-0085921.  
 PR 18-MAY-1998; 98US-0085922.  
 PR 18-MAY-1998; 98US-0085923.  
 PR 18-MAY-1998; 98US-0085924.  
 PR 18-MAY-1998; 98US-0085928.  
 PR 18-MAY-1998; 98US-0085925.  
 PR 18-MAY-1998; 98US-0085927.

XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Ruben SM, Florence K, Ni J, Rosen CA, Carter KC, Moore PA;  
 PI Olsen HS, Shi Y, Young PE, Wei F, Brewer LA, Soppet DR;  
 PI Lafleur DW, Endress GA, Ebner R;  
 XX WPI; 2000-062296/05.  
 DR P-PSDB; AAY76131.

XX New isolated human genes and the secreted polypeptides they encode,  
 PT useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders -  
 XX Claim 1; Page 300; 475pp; English.

XX AAZ65250 to AAZ65350 represent 97 isolated human secreted protein genes.  
 CC This sequence was found to be present on human chromosome 10.  
 CC AAY76124 to AAY76223 represent the secreted proteins encoded by the 97  
 CC human genes. The genes and their corresponding secreted polypeptides are  
 CC useful for preventing, treating or ameliorating medical conditions,  
 CC e.g. by protein or gene therapy. Also pathological conditions can be

CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new genes. Specific  
 CC uses are described for each of the 97 genes, based on which tissues they  
 CC are most highly expressed in, and include developing products for the  
 CC diagnosis or treatment of cancer, tumours, developmental abnormalities  
 CC and foetal deficiencies, blood disorders, diseases of the immune system,  
 CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive  
 CC disorders, schizophrenia, arthritis, psoriasis, sepsis, skin  
 CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney  
 CC disorders, digestive/endocrine disorders, infections and AIDS. The  
 CC polypeptides are also useful for identifying their binding partners.  
 CC The sequences shown in AAY76224 to AAY76424 represent fragments of the  
 CC secreted proteins.

XX SQ Sequence 1143 BP; 242 A; 339 C; 351 G; 210 T; 1 other;

Query Match 74.8%; Score 17.2; DB 21; Length 1143;  
 Best Local Similarity 86.4%; Pred. No. 94;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 CCAGGCGTCTCGTGGGCACAT 23  
 | ||||| ||||| |||||  
 Db 37 CGAGGCGTACGTGGGCCCAT 58

RESULT 20  
 ABL07567/c  
 ID ABL07567 standard; cDNA; 1851 BP.

XX AC ABL07567;  
 XX 26-MAR-2002 (first entry)  
 DT Drosophila melanogaster expressed polynucleotide SEQ ID NO 17183.

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 17183.  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 KW Drosophila melanogaster.  
 OS Drosophila melanogaster.  
 XX WO200171042-A2.  
 PN 27-SEP-2001.

XX PD 23-MAR-2001; 2001WO-US09231.  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 PA Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI; 2001-656860/75.  
 DR P-PSDB; ABB63464.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX Claim 1; SEQ ID NO 17183; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

```
XX SQ Sequence 1851 BP; 479 A; 533 C; 478 G; 361 T; 0 other;
Query Match 74.8%; Score 17.2; DB 23; Length 1851;
Best Local Similarity 86.4%; Pred. No. 97;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 CCAGCGCTCTCGTGGCCACAT 23
   ||| ||||| |||||
Db 1361 CCGGACGCTCGTGGCCACAT 1340

RESULT 21
ABL07566
ID ABL07566 standard; cDNA; 7832 BP.
XX
XX AC ABL07566;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 17180.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US09231.
XX
XX PR 23-MAR-2000; 2000US-191637P.
XX
XX PR 11-JUL-2000; 2000US-0614150.
XX
XX PA (PEKE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX DR P-PSDB; ABB63463.
XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX
XX PS Claim 1; SEQ ID NO 17180; 2lpp + Sequence Listing; English.
XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 7832 BP; 2131 A; 1618 C; 1789 G; 2294 T; 0 other;
Query Match 74.8%; Score 17.2; DB 23; Length 7832;
Best Local Similarity 86.4%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 CCAGCGCTCTCGTGGCCACAT 23
   ||| ||||| |||||
Db 2959 CCGGACGCTCTCGTGGCCACAT 2980

RESULT 22
ABK72958/c
ID ABK72958 standard; DNA; 1557 BP.
XX
XX AC ABK72958;
XX
XX DT 13-AUG-2002 (first entry)
XX
XX DE Bacillus licheniformis genomic sequence tag (GST) #249.
XX
XX KW Differential gene expression; genomic sequenced tag; GST;
XX KW altered culture condition; environmental stress;
XX KW physiological provocation; ds.
XX
XX OS Bacillus licheniformis.
XX PN WO200229113-A2.
XX
XX PD 11-APR-2002.
XX
XX PF 05-OCT-2001; 2001WO-US31437.
XX
XX PR 06-OCT-2000; 2000US-0680598.
XX
XX PR 27-MAR-2001; 2001US-279526P.
XX
XX PA (NOVO ) NOVOZYMES BIOTECH INC.
XX PA (NOVO ) NOVOZYMES AS.
XX
XX PI Berka R, Clausen IG;
XX
XX WPI; 2002-416684/44.
XX
XX PT Monitoring differential expression of several genes in first Bacillus
XX PT cell relative to expression of same genes in one or more second
XX PT Bacillus cells, by using substrate containing Bacillus genomic
XX PT sequenced tag array -
XX
XX PS Claim 4; SEQ ID NO 249; 200pp; English.
XX
XX CC The invention describes a method of monitoring differential expression of
XX CC genes in a first Bacillus cell relative to expression of the genes in
XX CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
XX CC isolated from Bacillus cells to a substrate containing array of Bacillus
XX CC genomic sequenced tags (GST), examining the array, and determining
XX CC relative gene expression by an observed hybridisation reporter signal of
XX CC a spot in the array. The method is useful for measuring the expression of
XX CC genes in a first Bacillus cell relative to expression of the same genes
XX CC in one or more second Bacillus cells. The method is useful for monitoring
XX CC global expression of several genes from a Bacillus cell, discovering new
XX CC genes, identifying possible functions of unknown open reading frames and
XX CC monitoring gene copy number variation and stability. Monitoring changes
XX CC in expression of genes may be used to provide a representation of the way
XX CC in which Bacillus cells adapt to changes in culture conditions,
XX CC environmental stress or other physiological provocation. Extensive
XX CC follow-up characterisation is unnecessary, when one spot on an array
XX CC equals one gene or one open reading frame, since sequence information is
XX CC available. This sequence represents a genomic sequence tag (GST) used in
XX CC the method of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 1557 BP; 406 A; 361 C; 417 G; 373 T; 0 other;
Query Match 73.0%; Score 16.8; DB 24; Length 1557;
Best Local Similarity 90.0%; Pred. NO. 1.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 4 AGCGCTCTCGTGGCCACAT 23
   ||||| ||| |||||
Db 284 AGCGCTCGGTGGCCACAT 265
```

```
RESULT 23
AAT35522
ID AAT35522 standard; DNA; 4042 BP.
XX
XX
AC AAT35522;
XX
XX 14-JAN-1997 (first entry)
XX
XX Human DNase I gene 5' flanking sequence.
XX
XX
XX
KW Gene targeting; gene activation; homologous recombination;
KW DNase I; cystic fibrosis; gene therapy; ds.
XX
XX Homo sapiens.
XX
XX WO96294111-A1.
XX
XX 26-SEP-1996.
XX
XX 12-MAR-1996; 96WO-US03377.
XX
XX 17-MAR-1995; 95US-0406030.
XX
XX (TRAN-) TRANSKARYOTIC THERAPIES INC.
XX
XX Hauge BM, Heartlein MW, Seiden RF, Treco DA;
XX
XX WPI; 1996-443186/44.
XX
XX Altering expression of genes encoding thrombopoietin, DNase I or
XX beta-interferon - using DNA constructs useful in gene therapy to
XX treat, e.g. cystic fibrosis and multiple sclerosis.
XX
XX Claim 27; Fig 10A-D; 115pp; English.
XX
XX A genomic DNA fragment (AAT35522) corresponds to nucleotides -4512
XX to -471 upstream of the initiation codon of the human DNase I
XX gene. It forms part of a DNA clone isolated from a human leukocyte
XX genomic DNA library; another portion of the clone, comprising part
XX of the 5' untranslated region and coding sequence of the TPO gene,
XX is given in AAT35523. Non-coding genomic sequences, e.g. exon 1,
XX within and upstream of the transcribed regions of the DNase I gene
XX may be used as targeting sequences in DNA constructs useful for
XX gene activation by homologous recombination. Novel genes are
XX generated in which exogenous and endogenous exons are operatively
XX linked. Expression of DNase I encoded by the endogenous gene can
XX be activated in the trachea and lungs of a cystic fibrosis patient,
XX thereby relieving respiratory distress.
XX
XX SQ Sequence 4042 BP; 971 A; 1081 C; 1110 G; 878 T; 2 other;
Query Match 73.08; Score 16.8; DB 17; Length 4042;
Best Local Similarity 90.08; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 CAGCGCTCTCGTGGGCCACA 22
||||| || |||||
Db 3550 CAGCGTGTCTGGGCCACA 3569

RESULT 24
AAL07078
ID AAL07078 standard; DNA; 25701 BP.
XX
XX
XX AAL07078;
XX
XX 21-NOV-2001 (first entry)
XX
XX Human reproductive system related antigen DNA SEQ ID NO: 9766.
XX
XX Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX
```

```
OS Homo sapiens.
XX
XX WO200155320-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184564.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226888.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.
XX 21-SEP-2000; 2000US-0234274.
XX 25-SEP-2000; 2000US-0234997.
XX 25-SEP-2000; 2000US-0234998.
XX 26-SEP-2000; 2000US-0235484.
```

PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239933.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PR (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-465570/50.  
XX Isolated nucleic acid molecule encoding a reproductive system antigen  
XX is used in preventing, treating or ameliorating a medical condition -  
PT Disclosure; SEQ ID NO 9766; 1297pp + Sequence Listing; English.  
XX  
PS The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a genomic sequence encoding a  
CC protein of the invention.  
XX  
SQ Sequence 25701 BP; 6452 A; 7118 C; 6915 G; 5216 T; 0 other;  
  
Query Match 73.0%; Score 16.8; DB 22; Length 25701;  
Best Local Similarity 90.0%; Pred. No. 1.8e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 3 CAGGCGTCTCGTGGCCACA 22  
||||||| || |||||||||  
Db 317 CAGGCGTGTCTGGCCACA 336  
  
RESULT 25  
AAL07077  
ID AAL07077 standard; DNA; 25758 BP.  
XX AC AAL07077;  
XX 21-NOV-2001 (first entry)  
DT  
XX Human reproductive system related antigen DNA SEQ ID NO: 9765.  
DE Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ds.  
XX Homo sapiens.  
XX PN WO200155320-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01339.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465570/50.  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen  
XX is used in preventing, treating or ameliorating a medical condition -  
XX  
XX Disclosure; SEQ ID NO 9765; 1297pp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
XX number of human reproductive system related antigens. These can be used  
XX in the prevention and treatment of reproductive system disorders,  
XX including cancer. The present sequence is a genomic sequence encoding a  
XX protein of the invention.  
XX  
XX SQ Sequence 25758 BP; 6456 A; 7140 C; 6936 G; 5226 T; 0 other;  
XX  
XX Query Match 73.0%; Score 16.8; DB 22; Length 25758;  
XX Best Local Similarity 90.0%; Pred. No. 1.8e+02;  
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Oy 3 CAGGCGTCTCGTGGGCACA 22  
Db 317 CAGGCGTGTCTGGGCACA 336  
|||||||  
|||||||

Search completed: December 2, 2002, 21:02:38  
Job time : 285 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2002, 20:56:18 ; Search time 50 seconds  
(without alignments)  
141.071 Million cell updates/sec

Title: US-09-700-906A-3

Perfect score: 23

Sequence: 1 accaggcgtctcgtgggcacat 23

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 239980

Minimum DB seq length: 66

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*

2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*

3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*

4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*

5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq:\*

6: /cgn2\_6/ptodata/1/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16.8	73.0	4042	4	US-08-406-030A-17
2	16.6	72.2	492	2	US-08-892-880-12
3	16.6	72.2	1112	3	US-08-369-822C-1
4	16.6	72.2	1112	3	US-08-582-776C-1
5	16.6	72.2	1112	3	US-08-434-831B-1
6	16.6	72.2	2029	4	US-09-232-160-13
7	16.6	72.2	2313	2	US-08-892-880-1
8	16.6	72.2	8910	3	US-08-369-822C-19
9	16.6	72.2	8910	3	US-08-779-764A-1
10	16.6	72.2	8910	3	US-08-582-776C-19
11	16.6	72.2	8910	3	US-08-434-831B-19
12	16.2	70.4	542	4	US-09-091-725-50
13	16.2	70.4	3157	6	5198347-3
14	16.2	70.4	6709	4	US-09-285-601-3
15	15.8	68.7	3098	4	US-09-232-200-58
16	15.8	68.7	3098	4	US-09-232-197-58
17	15.8	68.7	3098	4	US-09-232-201-58
18	15.6	67.8	175	4	US-09-060-756-708
19	15.6	67.8	335	1	US-08-171-718-116
20	15.6	67.8	335	3	US-08-478-087-116
21	15.6	67.8	1559	2	US-08-160-524A-1
22	15.6	67.8	1938	4	US-09-232-200-29
23	15.6	67.8	1938	4	US-09-232-197-29
24	15.6	67.8	1938	4	US-09-232-201-29
25	15.6	67.8	2983	4	US-08-982-785A-1
26	15.6	67.8	3217	4	US-09-232-200-64
27	15.6	67.8	3217	4	US-09-232-197-64

28	15.6	67.8	3217	4	US-09-232-201-64	Sequence 64, Appl
29	15.6	67.8	3841	3	US-08-888-077A-33	Sequence 33, Appl
30	15.4	67.0	3112	4	US-09-484-970B-146	Sequence 146, App
31	15.2	66.1	75	5	PCT-US92-01358-1	Sequence 1, Appli
32	15.2	66.1	162	1	US-08-275-916-1	Sequence 1, Appli
33	15.2	66.1	162	5	PCT-US93-03868-1	Sequence 1, Appli
34	15.2	66.1	423	4	US-09-641-638-566	Sequence 566, App
35	15.2	66.1	423	4	US-09-641-638-567	Sequence 567, App
36	15.2	66.1	423	4	US-09-641-638-568	Sequence 568, App
37	15.2	66.1	423	4	US-09-641-638-569	Sequence 569, App
38	15.2	66.1	498	4	US-08-998-416-180	Sequence 180, App
39	15.2	66.1	4260	2	US-08-658-665-38	Sequence 38, Appl
40	15.2	66.1	4260	4	US-08-796-101-2	Sequence 2, Appli
41	15.2	66.1	4260	4	US-09-085-273-38	Sequence 38, Appl
42	15.2	66.1	4465	4	US-09-485-636-1	Sequence 1, Appli
43	15.2	66.1	5234	2	US-08-658-665-73	Sequence 73, Appl
44	15.2	66.1	5234	4	US-08-796-101-37	Sequence 37, Appl
45	15.2	66.1	5234	4	US-09-085-273-73	Sequence 73, Appl

#### ALIGNMENTS

RESULT 1  
US-08-406-030A-17  
; Sequence 17, Application US/08406030A  
; Patent No. 6270989  
; GENERAL INFORMATION:  
; APPLICANT: Treco, Douglas A.  
; APPLICANT: Heartlein, Michael W.  
; APPLICANT: Hauge, Brian M.  
; APPLICANT: Seldgen, Richard F  
; TITLE OF INVENTION: Protein Production and Delivery  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.c.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/406.030A  
; FILING DATE: 17-MAR-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/243,391  
; FILING DATE: 13-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,586  
; FILING DATE: 03-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/911,533  
; FILING DATE: 10-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/787,840  
; FILING DATE: 05-NOV-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/789,188  
; FILING DATE: 05-NOV-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/11704  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/09627  
; FILING DATE: 05-NOV-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: TKT95-01  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4042 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-406-030A-17

Query Match 73.0%; Score 16.8; DB 4; Length 4042;

Best Local Similarity 90.0%; Pred. No. 29; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY 3 CAGGGCGTCTCGTGGCCACA 22

||||| ||| ||| ||| |||

Db 3550 CAGGGCGTCTCGTGGCCACA 3569

#### RESULT 2

US-08-892-880-12/c  
Sequence 12, Application US/08892880  
Patent No. 5942417

GENERAL INFORMATION:

APPLICANT: NI, JIAN  
APPLICANT: GENTZ, REINER L.  
APPLICANT: DILLON, PATRICK J.  
TITLE OF INVENTION: CD44-LIKE PROTEIN

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600

CITY: WASHINGTON

STATE: DC

COUNTRY: USA

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/892,880

FILING DATE: HEREWITH

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/021,762

FILING DATE: 15-JUL-1996

ATTORNEY/AGENT INFORMATION:

NAME: STEFFE, ERIC K

REGISTRATION NUMBER: 36,688

REFERENCE/DOCKET NUMBER: 1488.0490001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-2600

TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 492 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-892-880-12

Query Match 72.2%; Score 16.6; DB 2; Length 492;

Best Local Similarity 82.6%; Pred. No. 33; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 4; Indels 4; Gaps 0;

QY 1 ACCAGGCGTCTCGTGGCCACAT 23

Db 160 ACCAGGAGCCTCGTGGTCCAGAT 138

||||| | ||||| ||| ||

#### RESULT 3

US-08-369-822C-1/c

Sequence 1, Application US/08369822C

Patent No. 6015660

GENERAL INFORMATION:

APPLICANT: Lipkin, W. I.

APPLICANT: Briese, Thomas

APPLICANT: Kliche, Stefanie

APPLICANT: Schneider, Patrick A.

APPLICANT: Stitz, Lothar

APPLICANT: Schneemann, Anette

TITLE OF INVENTION: Borna Disease Viral Sequences,

TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous

TITLE OF INVENTION: System Diseases

NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fulbright & Jaworski, L.L.P.

STREET: 865 South Figueroa Street, 29th Floor

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90017-2571

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0

SOFTWARE: ASCII (DOS) TEXT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/369,822C

FILING DATE: 06-JAN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Churchill, Margaret A. (Ph.D.)

REGISTRATION NUMBER: 39,944

REFERENCE/DOCKET NUMBER: 1279-194XX

TELECOMMUNICATION INFORMATION:

TELEPHONE: 213/892-9200

TELEFAX: 213/680-4518

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1112 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-369-822C-1

Query Match 72.2%; Score 16.6; DB 3; Length 1112;

Best Local Similarity 82.6%; Pred. No. 34;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACCAGGCGTCTCGTGGCCACAT 23

||||| ||| ||| |||

Db 23 ACCAGGCGTCTCGTGGTGGCAT 1

#### RESULT 4

US-08-582-776C-1/c

Sequence 1, Application US/08582776C

Patent No. 6077510

GENERAL INFORMATION:

APPLICANT: Lipkin, W. I.

APPLICANT: Briese, Thomas

APPLICANT: Kliche, Stefanie

APPLICANT: Schneider, Patrick A.

APPLICANT: Stitz, Lothar

APPLICANT: Schneemann, Anette

;; TITLE OF INVENTION: Borna Disease Viral Sequences,  
;; TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
;; TITLE OF INVENTION: System Diseases  
;; NUMBER OF SEQUENCES: 61  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Fulbright & Jaworski, L.L.P.  
;; STREET: 865 South Figueroa Street, 29th Floor  
;; CITY: Los Angeles  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 90017-2576  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: WINDOWS NT  
;; SOFTWARE: ASCII DOS TEXT  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/582,776C  
;; FILING DATE: 04-JAN-1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/369,822  
;; FILING DATE: 06-JAN-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/434,831  
;; FILING DATE: 04-MAY-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Churchill, Margaret A.  
;; REGISTRATION NUMBER: 39,944  
;; REFERENCE/DOCKET NUMBER: 1279-194C2  
;; TELEPHONE: 213/892-9200  
;; TELEFAX: 213/680-4518  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1112 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cdna  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; US-08-582-776C-1

Query Match 72.2%; Score 16.6; DB 3; Length 1112;  
Best Local Similarity 82.6%; Pred. No. 34;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGGCACAT 23  
||||| ||||| ||||| ||||| |||||  
Db 23 ACCAGGCGTCTCTGGTGGCAT 1

RESULT 5  
US-08-434-831B-1/c  
; Sequence 1, Application US/08434831B  
; Patent No. 6113905  
; GENERAL INFORMATION:  
; APPLICANT: Lipkin, W. I.  
; APPLICANT: Briese, Stefan  
; APPLICANT: Kliche, Thomas  
; APPLICANT: Schneider, Patrick A.  
; APPLICANT: Stitz, Lothar  
; APPLICANT: Schneemann, Anette  
; TITLE OF INVENTION: Borna Disease Viral Sequences,  
; TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
; TITLE OF INVENTION: System Diseases  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski, L.L.P.  
; STREET: 865 South Figueroa Street, 29th Floor  
; CITY: Los Angeles  
; STATE: California

;; COUNTRY: USA  
;; ZIP: 90017-2571  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0  
;; SOFTWARE: ASCII (DOS) TEXT  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/434,831B  
;; FILING DATE: 04-MAY-1995  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/369,822  
;; FILING DATE: 06-JAN-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Churchill, Margaret A. (Ph.D.)  
;; REGISTRATION NUMBER: 39,944  
;; REFERENCE/DOCKET NUMBER: 1279-194C1  
;; TELEPHONE: 213/892-9200  
;; TELEFAX: 213/680-4518  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1112 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cdna to mRNA  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; US-08-434-831B-1

Query Match 72.2%; Score 16.6; DB 3; Length 1112;  
Best Local Similarity 82.6%; Pred. No. 34;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGGCACAT 23  
||||| ||||| ||||| ||||| |||||  
Db 23 ACCAGGCGTCTCTGGTGGCAT 1

RESULT 6  
US-09-232-160-13/c  
; Sequence 13, Application US/09232160  
; Patent No. 6368794  
; GENERAL INFORMATION:  
; APPLICANT: Steve Daniel  
; APPLICANT: James Gilmore  
; APPLICANT: Susan G. Stuart  
; APPLICANT: Laura Stuve  
; TITLE OF INVENTION: DETECTION OF ALTERED EXPRESSION OF GENES REGULATING CELL  
; TITLE OF INVENTION: PROLIFERATION  
; FILE REFERENCE: PA-0003 US  
; CURRENT APPLICATION NUMBER: US/09/232,160  
; CURRENT FILING DATE: 1999-01-15  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PERL Program  
; SEQ ID NO 13  
; LENGTH: 2029  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE: -  
; OTHER INFORMATION: 3044710  
; US-09-232-160-13

Query Match 72.2%; Score 16.6; DB 4; Length 2029;  
Best Local Similarity 82.6%; Pred. No. 35;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGGCACAT 23  
||||| ||||| ||||| ||||| |||||  
Db 244 ACCAGGAGCCTCGTGGTCCAGAT 222

```
RESULT 7
US-08-892-880-1/C
; Sequence 1, Application US/08892880
; Patent No. 5942417
; GENERAL INFORMATION:
; APPLICANT: NI, JIAN
; APPLICANT: GENTZ, REINER L.
; APPLICANT: DILLON, PATRICK J.
; TITLE OF INVENTION: CD44-LIKE PROTEIN
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/892.880
; FILING DATE: HERewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/021,762
; FILING DATE: 15-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0490001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2313 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 91..1056
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 154..1056
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 91..153
; US-08-892-880-1

Query Match 72.2%; Score 16.6; DB 2; Length 2313;
Best Local Similarity 82.6%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGCCACAT 23
||||| | ||||| ||| |
Db 152 ACCAGGAGCCTCGTGGTCCAGAT 130

RESULT 8
US-08-369-822C-19/c
; Sequence 19, Application US/08369822C
; Patent No. 6015660
; GENERAL INFORMATION:
; APPLICANT: Lipkin, W. I.
; APPLICANT: Briese, Thomas
; APPLICANT: Kliche, Stefanie
; APPLICANT: Schneider, Patrick A.

Query Match 72.2%; Score 16.6; DB 2; Length 2313;
Best Local Similarity 82.6%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGCCACAT 23
||||| | ||||| ||| |
Db 152 ACCAGGAGCCTCGTGGTCCAGAT 130

RESULT 9
US-08-779-764A-1/C
; Sequence 1, Application US/08779764A
; Patent No. 6057094
; GENERAL INFORMATION:
; APPLICANT: de la Torre, Juan C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SCREENING
; TITLE OF INVENTION: OF HUMAN BORNA DISEASE VIRUS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10550 No. 6057094th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: United States
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/779,764A
; FILING DATE: 16-DEC-1996

; APPLICANT: Stitz, Lothar
; APPLICANT: Schneemann, Anette
; TITLE OF INVENTION: Borna Disease Viral Sequences,
; TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski, L.L.P.
; STREET: 865 South Figueroa Street, 29th Floor
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017-2571
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0
; SOFTWARE: ASCII (DOS) TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/369,822C
; FILING DATE: 06-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Churchill, Margaret A. (Ph.D.)
; REGISTRATION NUMBER: 39,944
; REFERENCE/DOCKET NUMBER: 1279-194XX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213/892-9200
; TELEFAX: 213/680-4518
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8910 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to genomic RNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-369-822C-19

Query Match 72.2%; Score 16.6; DB 3; Length 8910;
Best Local Similarity 82.6%; Pred. No. 37;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGCCACAT 23
||||| | ||||| ||| |
Db 76 ACCAGGCGTCTCTGGGNGGCAT 54
```

; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fitting, Thomas  
; REGISTRATION NUMBER: 34,163  
; REFERENCE/DOCKET NUMBER: TSKI 465.0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 784-2937  
; TELEFAX: (619) 784-9399  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8910 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-779-764A-1

Query Match 72.2%; Score 16.6; DB 3; Length 8910;  
Best Local Similarity 82.6%; Pred. No. 37;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACCAGGCGTCTCGTGGCCACAT 23  
|||||  
Db 76 ACCAGGCGTCTCTGGTGGCAT 54

## RESULT 10

US-08-582-776C-19/c  
; Sequence 19, Application US/08582776C  
; Patent No. 6077510  
; GENERAL INFORMATION:  
; APPLICANT: Lipkin, W. I.  
; APPLICANT: Briese, Thomas  
; APPLICANT: Kliche, Stefanie  
; APPLICANT: Schneider, Patrick A.  
; APPLICANT: Stitz, Lothar  
; APPLICANT: Schneemann, Anette  
; TITLE OF INVENTION: Borna Disease Viral Sequences,  
; TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
; TITLE OF INVENTION: System Diseases  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski, L.L.P.  
; STREET: 865 South Figueroa Street, 29th Floor  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90017-2576  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: WINDOWS NT  
; SOFTWARE: ASCII DOS TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/582,776C  
; FILING DATE: 04-JAN-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/369,822  
; FILING DATE: 06-JAN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/434,831  
; FILING DATE: 04-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Churchill, Margaret A.  
; REGISTRATION NUMBER: 39,944  
; REFERENCE/DOCKET NUMBER: 1279-194C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 213/892-9200  
; TELEFAX: 213/680-4518  
; INFORMATION FOR SEQ ID NO: 19:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8910 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-582-776C-19

Query Match 72.2%; Score 16.6; DB 3; Length 8910;  
Best Local Similarity 82.6%; Pred. No. 37;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACCAGGCGTCTCGTGGCCACAT 23  
|||||  
Db 76 ACCAGGCGTCTCTGGTGGCAT 54

## RESULT 11

US-08-434-831B-19/c  
; Sequence 19, Application US/08434831B  
; Patent No. 6113905  
; GENERAL INFORMATION:  
; APPLICANT: Lipkin, W. I.  
; APPLICANT: Briese, Thomas  
; APPLICANT: Kliche, Stefanie  
; APPLICANT: Schneider, Patrick A.  
; APPLICANT: Stitz, Lothar  
; APPLICANT: Schneemann, Anette  
; TITLE OF INVENTION: Borna Disease Viral Sequences,  
; TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
; TITLE OF INVENTION: System Diseases  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski, L.L.P.  
; STREET: 865 South Figueroa Street, 29th Floor  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90017-2571  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0  
; SOFTWARE: ASCII (DOS) TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/434,831B  
; FILING DATE: 04-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/369,822  
; FILING DATE: 06-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Churchill, Margaret A. (Ph.D.)  
; REGISTRATION NUMBER: 39,944  
; REFERENCE/DOCKET NUMBER: 1279-194C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 213/892-9200  
; TELEFAX: 213/680-4518  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8910 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to genomic RNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-434-831B-19

Query Match 72.2%; Score 16.6; DB 3; Length 8910;  
Best Local Similarity 82.6%; Pred. No. 37;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGCGTCTCGTGGCCACAT 23  
|||||  
Db 76 ACCAGCGTCTCTGGTGCCAT 54

## RESULT 12

US-09-091-725-50/c  
; Sequence 50, Application US/09091725  
; Patent No. 6329141  
; GENERAL INFORMATION:

; APPLICANT:  
; TITLE OF INVENTION: Improved methods for transforming Phaffia  
; TITLE OF INVENTION: and recombinant DNA for use therein  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Morrison & Foerster llp  
; STREET: 2000 Pennsylvania Avenue, N.W.  
; CITY: Washington

; STATE: DC

; COUNTRY: United States of America  
; ZIP: 20006-1888

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/091.725

; FILING DATE: 23-DEC-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 95203620.0

; FILING DATE: 22-DEC-1995

; APPLICATION NUMBER: EP 96200943.7

; FILING DATE: 11-APR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: E. Victor Donahue

; REGISTRATION NUMBER: 35,492

; INFORMATION FOR SEQ ID NO: 50:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 542 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHEICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Phaffia rhodozyma

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 18..443

; OTHER INFORMATION: /product= "PRCDNA95"

US-09-091-725-50

Query Match 70.4%; Score 16.2; DB 4; Length 542;

Best Local Similarity 85.7%; Pred. No. 51;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CCAGCGTCTCGTGGCCACA 22

|||||

Db 85 CCAGGGTGGCGTGGCCACA 65

## RESULT 13

5198347-3

; Patent No. 5198347

; APPLICANT: MILLER, LOUIS H.; ADAMS, JOHN H.; KASLOW,

; DAVIC C.; FANG, XIANGDOUG

; TITLE OF INVENTION: DNA ENCODING PLASMODIUM VIVAX AND

; PLASMODIUM KNOWLESII DUFFY RECEPTOR

; NUMBER OF SEQUENCES: 27  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/554,837  
; FILING DATE: 20-JUL-1990  
; SEQ ID NO:3;  
; LENGTH: 3157  
; 5198347-3

Query Match 70.4%; Score 16.2; DB 6; Length 3157;

Best Local Similarity 85.7%; Pred. No. 54;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CAGCGTCTCGTGGCCACAT 23

|||||

Db 1248 CAGGGTCACTTGGCCACAT 1268

## RESULT 14

US-09-285-601-3

; Sequence 3, Application US/09285601

; Patent No. 6248528

; GENERAL INFORMATION:

; APPLICANT: Chen, Hong

; APPLICANT: Freimer, Nelson

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND

; TITLE OF INVENTION: TREATMENT OF NEUROPSYCHIATRIC DISORDERS

; FILE REFERENCE: 7853-089

; CURRENT APPLICATION NUMBER: US/09/285,601

; CURRENT FILING DATE: 1999-04-02

; EARLIER APPLICATION NUMBER: 60/080,841

; EARLIER FILING DATE: 1998-04-06

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 6709

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-285-601-3

Query Match 70.4%; Score 16.2; DB 4; Length 6709;

Best Local Similarity 85.7%; Pred. No. 56;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CAGCGTCTCGTGGCCACAT 23

|||||

Db 72 CAGGCCCTCTCTGGCCCTCAT 92

## RESULT 15

US-09-232-200-58

; Sequence 58, Application US/09232200A

; Patent No. 6288213

; GENERAL INFORMATION:

; APPLICANT: Stahl, Andreas

; APPLICANT: Hirsch, David J.

; APPLICANT: Lodish, Harvey F.

; APPLICANT: Gimeno, Ruth E.

; APPLICANT: Tartaglia, Louis A.

; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS

; FILE REFERENCE: WHI97-21p3MB

; CURRENT APPLICATION NUMBER: US/09/232,200A

; CURRENT FILING DATE: 1999-01-14

; EARLIER APPLICATION NUMBER: 60/071,374

; EARLIER FILING DATE: 1998-01-15

; EARLIER APPLICATION NUMBER: 60/093,491

; EARLIER FILING DATE: 1998-07-20

; EARLIER APPLICATION NUMBER: 60/110,941

; EARLIER FILING DATE: 1998-12-04

; NUMBER OF SEQ ID NOS: 105

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 58

; LENGTH: 3098

; TYPE: DNA

```
; ORGANISM: Rattus norvegicus
US-09-232-200-58

Query Match      68.7%; Score 15.8; DB 4; Length 3098;
Best Local Similarity 89.5%; Pred. No. 83;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 GCGCTCTCGTGGCCACAT 23
   ||| ||||| ||||| ||
Db 1417 GCCTTCTCGTGGCCAGAT 1435

RESULT 16
US-09-232-197-58
; Sequence 58, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 58
; LENGTH: 3098
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-232-197-58

Query Match      68.7%; Score 15.8; DB 4; Length 3098;
Best Local Similarity 89.5%; Pred. No. 83;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 GCGCTCTCGTGGCCACAT 23
   ||| ||||| ||||| ||
Db 1417 GCCTTCTCGTGGCCAGAT 1435

RESULT 17
US-09-232-201-58
; Sequence 58, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MC
; CURRENT APPLICATION NUMBER: US/09/232,201A
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 58
; LENGTH: 3098
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-232-201-58

Query Match      68.7%; Score 15.8; DB 4; Length 3098;
Best Local Similarity 89.5%; Pred. No. 83;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 GCGCTCTCGTGGCCACAT 23
   ||| ||||| ||||| ||
Db 1417 GCCTTCTCGTGGCCAGAT 1435

RESULT 18
US-09-060-756-708
; Sequence 708, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/09/060,756
; CURRENT FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 708
; LENGTH: 175
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (various positions within the sequence)
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-708

Query Match      67.8%; Score 15.6; DB 4; Length 175;
Best Local Similarity 81.8%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CCAGGCGTCTCGTGGCCACAT 23
   ||| | |||| | ||||| ||
Db 95 CCAGGTGGCTCGTGGCCACAT 116

RESULT 19
US-08-171-718-116
; Sequence 116, Application US/08171718
; Patent No. 5707863
; GENERAL INFORMATION:
; APPLICANT: Trofatter, James A.
; APPLICANT: MacCollin, Mia M.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: Tumor Suppressor Gene Merlin and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/171,718
; FILING DATE: 22-DEC-1993
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/108,808
; FILING DATE: 19-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/022,034
; FILING DATE: 25-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/026,063
; FILING DATE: 04-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Anne
; REGISTRATION NUMBER: 36,463
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2500
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 116:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 335 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; US-08-171-718-116

Query Match      67.8%; Score 15.6; DB 1; Length 335;
Best Local Similarity 81.8%; Pred. No. 94;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGGCCACA 22
   |||| || || |||||
Db 46 ACCAAGCAGCTTGTGGGCCACA 67

RESULT 20
US-08-478-087-116
; Sequence 116, Application US/08478087
; Patent No. 6077685
; GENERAL INFORMATION:
; APPLICANT: Trofatter, James A.
; APPLICANT: MacCollin, Mia M.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: Tumor Suppressor Gene Merlin and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,087
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/171,718
; FILING DATE: 22-DEC-1993
; APPLICATION NUMBER: US 08/108,808
; FILING DATE: 19-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/022,034
; FILING DATE: 25-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/026,063
```

```
; FILING DATE: 04-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Anne
; REGISTRATION NUMBER: 36,463
; REFERENCE/DOCKET NUMBER: 0609.3850003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 116:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 335 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; US-08-478-087-116

Query Match      67.8%; Score 15.6; DB 3; Length 335;
Best Local Similarity 81.8%; Pred. No. 94;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGGCCACA 22
   |||| || || |||||
Db 46 ACCAAGCAGCTTGTGGGCCACA 67

RESULT 21
US-08-160-524A-1/c
; Sequence 1, Application US/08160524A
; Patent No. 5851761
; GENERAL INFORMATION:
; APPLICANT: McAdam, Ruth Anne
; APPLICANT: Dale, Jeremy W.
; APPLICANT: Zainuddin, Zainul Fadziruddin B.
; APPLICANT: Catty, David
; TITLE OF INVENTION: PROBES, KITS AND METHODS FOR THE
; TITLE OF INVENTION: DETECTION AND DIFFERENTIATION OF MYCOBACTERIA
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert,
; ADDRESSEE: Attn: Walter H. Dreger
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/160,524A
; FILING DATE: 01-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/752,661
; FILING DATE: 18-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8903968.9
; FILING DATE: 22-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9000411.0
; FILING DATE: 09-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB90/00276
; FILING DATE: 22-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Walter H.
; REGISTRATION NUMBER: 24,190
; REFERENCE/DOCKET NUMBER: A-55387-1/WH
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
```



; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1559 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..853  
US-08-160-524A-1

Query Match 67.8%; Score 15.6; DB 2; Length 1559;  
Best Local Similarity 81.8%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CCAGGCGTCTCGTGGCCACAT 23  
||| | ||||| |||||  
Db 1271 CGAGGTGGCTCGTGGCCACAT 1250

RESULT 22  
US-09-232-200-29  
; Sequence 29, Application US/09232200A  
; Patent No. 6288213  
; GENERAL INFORMATION:  
; APPLICANT: Stahl, Andreas  
; APPLICANT: Hirsch, David J.  
; APPLICANT: Lodish, Harvey F.  
; APPLICANT: Gimeno, Ruth E.  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
; FILE REFERENCE: WHI97-21p3MB  
; CURRENT APPLICATION NUMBER: US/09/232,200A  
; CURRENT FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: 60/071,374  
; EARLIER FILING DATE: 1998-01-15  
; EARLIER APPLICATION NUMBER: 60/093,491  
; EARLIER FILING DATE: 1998-07-20  
; EARLIER APPLICATION NUMBER: 60/110,941  
; EARLIER FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 29  
; LENGTH: 1938  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-232-200-29

Query Match 67.8%; Score 15.6; DB 4; Length 1938;  
Best Local Similarity 81.8%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CCAGGCGTCTCGTGGCCACAT 23  
||| | ||||| |||||  
Db 1340 CCGGCTTCTCGTGGCCAGAT 1361

RESULT 23  
US-09-232-197-29  
; Sequence 29, Application US/09232197A  
; Patent No. 6300096  
; GENERAL INFORMATION:  
; APPLICANT: Stahl, Andreas  
; APPLICANT: Hirsch, David J.  
; APPLICANT: Lodish, Harvey F.  
; APPLICANT: Gimeno, Ruth E.  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
; FILE REFERENCE: WHI97-21p3MA  
; CURRENT APPLICATION NUMBER: US/09/232,197A  
; CURRENT FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: 60/071,374

; EARLIER FILING DATE: 1998-01-15  
; EARLIER APPLICATION NUMBER: 60/093,491  
; EARLIER FILING DATE: 1998-07-20  
; EARLIER APPLICATION NUMBER: 60/110,941  
; EARLIER FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 29  
; LENGTH: 1938  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-232-197-29

Query Match 67.8%; Score 15.6; DB 4; Length 1938;  
Best Local Similarity 81.8%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CCAGGCGTCTCGTGGCCACAT 23  
||| | ||||| |||||  
Db 1340 CCGGCTTCTCGTGGCCAGAT 1361

RESULT 24  
US-09-232-201-29  
; Sequence 29, Application US/09232201A  
; Patent No. 6348321  
; GENERAL INFORMATION:  
; APPLICANT: Stahl, Andreas  
; APPLICANT: Hirsch, David J.  
; APPLICANT: Lodish, Harvey F.  
; APPLICANT: Gimeno, Ruth E.  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
; FILE REFERENCE: WHI97-21p3MC  
; CURRENT APPLICATION NUMBER: US/09/232,201A  
; CURRENT FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: 60/071,374  
; EARLIER FILING DATE: 1998-01-15  
; EARLIER APPLICATION NUMBER: 60/093,491  
; EARLIER FILING DATE: 1998-07-20  
; EARLIER APPLICATION NUMBER: 60/110,941  
; EARLIER FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 29  
; LENGTH: 1938  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-232-201-29

Query Match 67.8%; Score 15.6; DB 4; Length 1938;  
Best Local Similarity 81.8%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CCAGGCGTCTCGTGGCCACAT 23  
||| | ||||| |||||  
Db 1340 CCGGCTTCTCGTGGCCAGAT 1361

RESULT 25  
US-08-982-785A-1/c  
; Sequence 1, Application US/08982785A  
; Patent No. 6258929  
; GENERAL INFORMATION:  
; APPLICANT: Kosik, Kenneth S.  
; APPLICANT: Zhou, Jianhua  
; TITLE OF INVENTION: ALARM RELATED PEPTIDES AND  
; TITLE OF INVENTION: NUCLEIC ACIDS AND DIAGNOSIS USING THEM  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston

STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/982,785A  
FILING DATE: 02-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/031,556  
FILING DATE: 02-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 05311/018001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2983 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 366...2633  
US-08-982-785A-1

Query Match 67.8%; Score 15.6; DB 4; Length 2983;  
Best Local Similarity 81.8%; Pred. No. Ie+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGGCACA 22  
||||||| | ||||| |||  
Db 163 ACCAGGCGTGTGGTGGGCACA 142

Search completed: December 2, 2002, 21:03:29  
Job time : 54 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2002, 21:02:48 ; Search time 54 seconds  
(without alignments)  
164.031 Million cell updates/sec

Title: US-09-700-906A-3

Perfect score: 23

Sequence: 1 accaggcgctctcggtggccacat 23

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 498988

Minimum DB seq length: 66

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCT05\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	23	100.0	10432	10	US-09-919-172-97
c 2	17.2	74.8	189	10	Sequence 97, Appl
c 3	17.2	74.8	573	10	Sequence 25850, A
c 4	16.8	73.0	1057	10	Sequence 9296, Ap
c 5	16.6	72.2	2372	9	Sequence 249, App
c 6	16.6	72.2	2372	10	Sequence 200, App
c 7	16.6	72.2	2372	10	Sequence 200, App
c 8	16.6	72.2	2372	12	Sequence 200, App
c 9	16.6	72.2	2404	10	Sequence 5, Appli
c 10	16.6	72.2	2404	10	Sequence 849, App
c 11	16.4	71.3	339	10	Sequence 853, App
c 12	16.2	70.4	10378	10	Sequence 3049, Ap
c 13	16.2	70.4	10472	10	Sequence 1616, Ap
c 14	16	69.6	10972	10	Sequence 1615, Ap
c 15	16	69.6	10957	10	Sequence 555, App
c 16	16	69.6	13217	10	Sequence 903, App
c 17	16	69.6	13217	10	Sequence 556, App
c 18	16	69.6	13220	10	Sequence 904, App
c 19	16	69.6	13220	10	Sequence 554, App
c 20	16	69.6	13220	10	Sequence 902, App

c 20	15.6	67.8	305	10	US-09-864-761-18859
c 21	15.6	67.8	345	10	US-09-815-343-1122
c 22	15.6	67.8	475	10	US-09-864-761-2116
c 23	15.6	67.8	476	10	US-09-864-761-6316
c 24	15.6	67.8	921	10	US-09-974-300-4778
c 25	15.6	67.8	2685	10	US-09-815-242-5994
c 26	15.4	67.0	1702	10	US-09-908-711-64
c 27	15.4	67.0	1702	10	US-09-908-711-64
c 28	15.4	67.0	2858	10	US-09-764-853-378
c 29	15.4	67.0	2858	10	US-09-908-711-18
c 30	15.4	67.0	3007	10	US-09-764-853-190
c 31	15.4	67.0	3111	10	US-09-799-777-135
c 32	15.2	66.1	517	9	US-09-349-015-8
c 33	15.2	66.1	2120	10	US-10-040-739-209
c 34	15.2	66.1	4465	10	US-09-801-574-29
c 35	15.2	66.1	4465	12	US-09-964-824A-267
c 36	15.2	66.1	32190	10	US-10-044-090-63
c 37	15	65.2	553	10	US-09-764-887-338
c 38	15	65.2	1072	10	US-09-864-761-7076
c 39	15	65.2	1083	10	US-09-737-149-31
c 40	15	65.2	1094	10	US-09-737-149-32
c 41	15	65.2	1120	10	US-09-737-149-38
c 42	15	65.2	1685	9	US-09-737-149-36
c 43	15	65.2	1685	9	US-09-954-531-133
c 44	15	65.2	1685	9	US-09-954-531-354
c 45	15	65.2	2296	10	US-09-962-436-261
c 46	15	65.2	2296	10	US-09-822-849A-259

ALIGNMENTS

RESULT 1

US-09-919-172-97/c  
; Sequence 97, Application US/09919172  
; Patent No. US20020119463A1  
; GENERAL INFORMATION:  
; APPLICANT: Paris, Mary  
; APPLICANT: Turner, Christopher M.  
; TITLE OF INVENTION: PROSTATE CANCER MARKERS  
; FILE REFERENCE: PA-0036 US  
; CURRENT APPLICATION NUMBER: US/09/919,172  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: 60/222,469  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 102  
; SOFTWARE: PERL Program  
; SEQ ID NO 97  
; LENGTH: 10432  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20020119463A1 2700132CBI  
US-09-919-172-97

Query Match 100.0%; Score 23; DB 10; Length 10432;  
Best Local Similarity 100.0%; Pred. No. 0.031;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGCCACAT 23  
|||||  
Db 150 ACCAGGCGTCTCGTGGCCACAT 128

RESULT 2

US-09-864-761-25850  
; Sequence 25850, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng

;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

;; FILE REFERENCE: Aecmica-X-1

;; CURRENT APPLICATION NUMBER: US 09/864,761

;; PRIOR FILING DATE: 2001-05-23

;; PRIOR APPLICATION NUMBER: US 60/180,312

;; PRIOR FILING DATE: 2000-02-04

;; PRIOR APPLICATION NUMBER: US 60/207,456

;; PRIOR FILING DATE: 2000-05-26

;; PRIOR APPLICATION NUMBER: US 09/632,366

;; PRIOR FILING DATE: 2000-08-03

;; PRIOR APPLICATION NUMBER: GB 24263.6

;; PRIOR FILING DATE: 2000-10-04

;; PRIOR APPLICATION NUMBER: US 60/236,359

;; PRIOR FILING DATE: 2000-09-27

;; PRIOR APPLICATION NUMBER: PCT/US01/00666

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00667

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00664

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00669

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00665

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00668

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00663

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00662

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00661

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00670

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: US 60/234,687

;; PRIOR FILING DATE: 2000-09-21

;; PRIOR APPLICATION NUMBER: US 09/608,408

;; PRIOR FILING DATE: 2000-06-30

;; PRIOR APPLICATION NUMBER: US 09/774,203

;; PRIOR FILING DATE: 2001-01-29

;; NUMBER OF SEQ ID NOS: 49117

;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

;; SEQ ID NO 25850

;; LENGTH: 189

;; TYPE: DNA

;; ORGANISM: Homo sapiens

;; FEATURE:

;; OTHER INFORMATION: MAP TO AC008151.1

;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5

;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.4

;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.2

;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.8

;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3

;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.9

;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.7

;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.4

;; OTHER INFORMATION: NT HTT: gi6677874, EVALUE 4.40e-01

;; OTHER INFORMATION: EST\_HUMAN HTT: A1580035.1, EVALUE 2.60e+00

;; OTHER INFORMATION: SWISSPROT HTT: Q64411, EVALUE 1.00e+00

US-09-864-761-25850

Query Match 74.8%; Score 17.2; DB 10; Length 189;

Best Local Similarity 86.4%; Pred. No. 15;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGGCCACA 22

||||| ||||| ||||| ||||| |||||

Db 1 ACCATGGGTCACGTGGGCCACA 22

||||| ||||| ||||| ||||| |||||

RESULT 3

US-09-864-761-9296

;; Sequence 9296, Application US/09864761

;; Patent NO. US20020048763A1

;; GENERAL INFORMATION:

;; APPLICANT: Penn, Sharon G.

;; APPLICANT: Rank, David R.

;; APPLICANT: Hanzel, David K.

;; APPLICANT: Chen, Wensheng

;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

;; FILE REFERENCE: Aecmica-X-1

;; CURRENT APPLICATION NUMBER: US/09/864,761

;; CURRENT FILING DATE: 2001-05-23

;; PRIOR APPLICATION NUMBER: US 60/180,312

;; PRIOR FILING DATE: 2000-02-04

;; PRIOR APPLICATION NUMBER: US 60/207,456

;; PRIOR FILING DATE: 2000-05-26

;; PRIOR APPLICATION NUMBER: US 09/632,366

;; PRIOR FILING DATE: 2000-08-03

;; PRIOR APPLICATION NUMBER: GB 24263.6

;; PRIOR FILING DATE: 2000-10-04

;; PRIOR APPLICATION NUMBER: US 60/236,359

;; PRIOR FILING DATE: 2000-09-27

;; PRIOR APPLICATION NUMBER: PCT/US01/00666

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00667

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00664

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00669

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00665

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00668

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00663

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00662

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00661

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00670

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: US 60/234,687

;; PRIOR FILING DATE: 2000-09-21

;; PRIOR APPLICATION NUMBER: US 09/608,408

;; PRIOR FILING DATE: 2000-06-30

;; PRIOR APPLICATION NUMBER: US 09/774,203

;; PRIOR FILING DATE: 2001-01-29

;; NUMBER OF SEQ ID NOS: 49117

;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

;; SEQ ID NO 9296

;; LENGTH: 573

;; TYPE: DNA

;; ORGANISM: Homo sapiens

;; FEATURE:

;; OTHER INFORMATION: MAP TO AC008151.1

;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5

;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.4

;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.2

;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.8

;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3

;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.9

;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.7

;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.4

;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.4

US-09-864-761-9296

Query Match 74.8%; Score 17.2; DB 10; Length 573;

Best Local Similarity 86.4%; Pred. No. 16;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGGCCACA 22

||||| ||||| ||||| ||||| |||||

Db 267 ACCATGGGTCACGTGGGCCACA 288

||||| ||||| ||||| ||||| |||||

RESULT 4  
US-09-974-300-249/c  
; Sequence 249, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: 10085-500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 249  
; LENGTH: 1557  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-249

Query Match 73.0%; Score 16.8; DB 10; Length 1557;  
Best Local Similarity 90.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 AGCGTCGTCGTGGGCCACAT 23  
||||||| ||| |||||||  
Db 284 AGCGTCGTCGTGGGCCACAT 265

RESULT 5  
US-09-905-291A-200/c  
; Sequence 200, Application US/09905291A  
; Patent No. US20020160374A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnovers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/905,291A  
; CURRENT FILING DATE: 2001-07-12  
; PRIOR APPLICATION NUMBER: PCT/US00/04414  
; PRIOR FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: US 60/143,048

; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 60/145,698  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: US 60/146,222  
; PRIOR FILING DATE: 1999-07-28  
; PRIOR APPLICATION NUMBER: PCT/US99/20594  
; PRIOR FILING DATE: 1999-09-08  
; PRIOR APPLICATION NUMBER: PCT/US99/20944  
; PRIOR FILING DATE: 1999-09-13  
; PRIOR APPLICATION NUMBER: PCT/US99/21090  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/21547  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/23089  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: PCT/US99/28214  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: PCT/US99/28313  
; PRIOR FILING DATE: 1999-11-30  
; PRIOR APPLICATION NUMBER: PCT/US99/28564  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/28565  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/30095  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: PCT/US99/30911  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US99/30999  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US00/00219  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 423  
; SEQ ID NO 200  
; LENGTH: 2372  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-905-291A-200

Query Match 72.2%; Score 16.6; DB 9; Length 2372;  
Best Local Similarity 82.6%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCGTCGTGGGCCACAT 23  
||||||| ||||||| ||| ||  
Db 221 ACCAGGAGCCTCGTGTCCAGAT 199

RESULT 6  
US-09-909-320-200/c  
; Sequence 200, Application US/09909320  
; Patent No. US20020132240A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnovers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann

; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/909,320  
; PRIOR FILING DATE: 2002-01-04  
; PRIOR APPLICATION NUMBER: PCT/US00/04414  
; PRIOR FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: US 60/143,048  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 60/145,698  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: US 60/146,222  
; PRIOR FILING DATE: 1999-07-28  
; PRIOR APPLICATION NUMBER: PCT/US99/20594  
; PRIOR FILING DATE: 1999-09-08  
; PRIOR APPLICATION NUMBER: PCT/US99/20944  
; PRIOR FILING DATE: 1999-09-13  
; PRIOR APPLICATION NUMBER: PCT/US99/21090  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/21547  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/23089  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: PCT/US99/28214  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: PCT/US99/28313  
; PRIOR FILING DATE: 1999-11-30  
; PRIOR APPLICATION NUMBER: PCT/US99/28564  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/28565  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/30095  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: PCT/US99/30911  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US99/30999  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US00/00219  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 423  
; SEQ ID NO 200  
; LENGTH: 2372  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; ORGANISM: Homo sapiens  
US-09-909-320-200

Query Match 72.2%; Score 16.6; DB 10; Length 2372;  
Best Local Similarity 82.6%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCGTGGGCACAT 23  
||||| | ||||| ||| ||  
Db 221 ACCAGGAGCCTCGTGTCCAGAT 199

RESULT 7  
US-09-909-088B-200/c  
; Sequence 200, Application US/09909088B  
; Patent No. US20020146709A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnovers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/909,088B  
; CURRENT FILING DATE: 2001-07-18  
; PRIOR APPLICATION NUMBER: PCT/US00/04414  
; PRIOR FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: US 60/143,048  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 60/145,698  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: US 60/146,222  
; PRIOR FILING DATE: 1999-07-28  
; PRIOR APPLICATION NUMBER: PCT/US99/20594  
; PRIOR FILING DATE: 1999-09-08  
; PRIOR APPLICATION NUMBER: PCT/US99/20944  
; PRIOR FILING DATE: 1999-09-13  
; PRIOR APPLICATION NUMBER: PCT/US99/21090  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/21547  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/23089  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: PCT/US99/28214  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: PCT/US99/28313  
; PRIOR FILING DATE: 1999-11-30  
; PRIOR APPLICATION NUMBER: PCT/US99/28564  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/28565  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/30095  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: PCT/US99/30911  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US99/30999  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US00/00219  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 423  
; SEQ ID NO 200  
; LENGTH: 2372  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; ORGANISM: Homo sapiens  
US-09-909-088B-200

Query Match 72.2%; Score 16.6; DB 10; Length 2372;  
Best Local Similarity 82.6%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCGTGGGCACAT 23  
||||| | ||||| ||| ||  
Db 221 ACCAGGAGCCTCGTGTCCAGAT 199

RESULT 8  
US-10-006-867-5/c

; Sequence 5, Application US/10006867  
; Patent No. US20020119130A1  
; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Collin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/006,867  
; CURRENT FILING DATE: 2001-12-06  
; PRIOR APPLICATION NUMBER: 60/063435  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/064215  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/082797  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/083495  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/085579  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/087759  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/088021  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088029  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088030  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088734  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088740  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088811  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088824  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088825  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088863  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/089105  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089514  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089653  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089952  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: 60/090246  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090444  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090688  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090696  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090862  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: 60/091628  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/096012  
; PRIOR FILING DATE: 1998-08-10  
; PRIOR APPLICATION NUMBER: 60/096757  
; PRIOR FILING DATE: 1998-08-17  
; PRIOR APPLICATION NUMBER: 60/096949  
; PRIOR FILING DATE: 1998-08-18  
; PRIOR APPLICATION NUMBER: 60/096959  
; PRIOR FILING DATE: 1998-08-18  
; PRIOR APPLICATION NUMBER: 60/097954  
; PRIOR FILING DATE: 1998-08-26  
; PRIOR APPLICATION NUMBER: 60/097971  
; PRIOR FILING DATE: 1998-08-26  
; PRIOR APPLICATION NUMBER: 60/097979  
; PRIOR FILING DATE: 1998-08-26  
; PRIOR APPLICATION NUMBER: 60/098749  
; PRIOR FILING DATE: 1998-09-01  
; PRIOR APPLICATION NUMBER: 60/099741  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 60/099763  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 60/099792  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 60/099812  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 60/099815  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 60/100627  
; PRIOR FILING DATE: 1998-09-16  
; PRIOR APPLICATION NUMBER: 60/100662  
; PRIOR FILING DATE: 1998-09-16  
; PRIOR APPLICATION NUMBER: 60/100683  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: 60/100684  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: 60/100930  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: 60/101279  
; PRIOR FILING DATE: 1998-09-22  
; PRIOR APPLICATION NUMBER: 60/101475  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101738  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/101743  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/101916  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/102570  
; PRIOR FILING DATE: 1998-09-30  
; PRIOR APPLICATION NUMBER: 60/103449  
; PRIOR FILING DATE: 1998-10-06  
; PRIOR APPLICATION NUMBER: 60/103678  
; PRIOR FILING DATE: 1998-10-08  
; PRIOR APPLICATION NUMBER: 60/103679  
; PRIOR FILING DATE: 1998-10-08  
; PRIOR APPLICATION NUMBER: 60/103711  
; PRIOR FILING DATE: 1998-10-08  
; PRIOR APPLICATION NUMBER: 60/105000  
; PRIOR FILING DATE: 1998-10-20  
; PRIOR APPLICATION NUMBER: 60/105002  
; PRIOR FILING DATE: 1998-10-20  
; PRIOR APPLICATION NUMBER: 60/105881  
; PRIOR FILING DATE: 1998-10-27  
; PRIOR APPLICATION NUMBER: 60/106030  
; PRIOR FILING DATE: 1998-10-28  
; PRIOR APPLICATION NUMBER: 60/106464  
; PRIOR FILING DATE: 1998-10-30  
; PRIOR APPLICATION NUMBER: 60/106856  
; PRIOR FILING DATE: 1998-11-03  
; PRIOR APPLICATION NUMBER: 60/108807  
; PRIOR FILING DATE: 1998-11-17  
; PRIOR APPLICATION NUMBER: 60/112419  
; PRIOR FILING DATE: 1998-12-15  
; PRIOR APPLICATION NUMBER: 60/112422  
; PRIOR FILING DATE: 1998-12-15  
; PRIOR APPLICATION NUMBER: 60/112853  
; PRIOR FILING DATE: 1998-12-16  
; PRIOR APPLICATION NUMBER: 60/113011  
; PRIOR FILING DATE: 1998-12-16  
; PRIOR APPLICATION NUMBER: 60/112854

; PRIOR FILING DATE: 1998-12-16  
; PRIOR APPLICATION NUMBER: 60/113300  
; PRIOR FILING DATE: 1998-12-22  
; PRIOR APPLICATION NUMBER: 60/113408  
; PRIOR FILING DATE: 1998-12-22  
; PRIOR APPLICATION NUMBER: 60/113430  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113621  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/114223  
; PRIOR FILING DATE: 1998-12-30  
; PRIOR APPLICATION NUMBER: 60/115614  
; PRIOR FILING DATE: 1999-01-12  
; PRIOR APPLICATION NUMBER: 60/116527  
; PRIOR FILING DATE: 1999-01-20  
; PRIOR APPLICATION NUMBER: 60/116843  
; PRIOR FILING DATE: 1999-01-22  
; PRIOR APPLICATION NUMBER: 60/119285  
; PRIOR FILING DATE: 1999-02-09  
; PRIOR APPLICATION NUMBER: 60/119287  
; PRIOR FILING DATE: 1999-02-09  
; PRIOR APPLICATION NUMBER: 60/119525  
; PRIOR FILING DATE: 1999-02-10  
; PRIOR APPLICATION NUMBER: 60/119549  
; PRIOR FILING DATE: 1999-02-10  
; PRIOR APPLICATION NUMBER: 60/120014  
; PRIOR FILING DATE: 1999-02-11  
; PRIOR APPLICATION NUMBER: 60/129122  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: 60/129674  
; PRIOR FILING DATE: 1999-04-16  
; PRIOR APPLICATION NUMBER: 60/131291  
; PRIOR FILING DATE: 1999-04-27  
; PRIOR APPLICATION NUMBER: 60/138387  
; PRIOR FILING DATE: 1999-06-09  
; PRIOR APPLICATION NUMBER: 60/144791  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/169495  
; PRIOR FILING DATE: 1999-12-07  
; PRIOR APPLICATION NUMBER: 60/175481  
; PRIOR FILING DATE: 2000-01-11  
; PRIOR APPLICATION NUMBER: 60/191007  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/193997  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/380139  
; PRIOR FILING DATE: 1998-08-25  
; PRIOR APPLICATION NUMBER: 09/311832  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 09/380137  
; PRIOR FILING DATE: 1999-08-25  
; PRIOR APPLICATION NUMBER: 09/380138  
; PRIOR FILING DATE: 1999-08-25  
; PRIOR APPLICATION NUMBER: 09/380142

Query Match 72.2%; Score 16.6; DB 12; Length 2372;  
Best Local Similarity 82.6%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGCCACAT 23  
||||| | ||||| | |||  
Db 221 ACCAGGAGCCTCGTGGTCCAGAT 199

RESULT 9  
US-09-833-381-849/c  
; Sequence 849, Application US/09833381  
; Patent No. US20020132090A1  
; GENERAL INFORMATION:  
; APPLICANT: Robison, Keith E.  
; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs  
; FILE REFERENCE: 5800-119  
; CURRENT APPLICATION NUMBER: US/09/833,381

; CURRENT FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 09/516,448  
; PRIOR FILING DATE: 2000-02-29  
; NUMBER OF SEQ ID NOS: 2050  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 849  
; LENGTH: 2404  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-833-381-849

Query Match 72.2%; Score 16.6; DB 10; Length 2404;  
Best Local Similarity 82.6%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGCCACAT 23  
||||| | ||||| | |||  
Db 251 ACCAGGAGCCTCGTGGTCCAGAT 229

RESULT 10  
US-09-833-381-853/c  
; Sequence 853, Application US/09833381  
; Patent No. US20020132090A1  
; GENERAL INFORMATION:  
; APPLICANT: Robison, Keith E.  
; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs  
; FILE REFERENCE: 5800-119  
; CURRENT APPLICATION NUMBER: US/09/833,381  
; CURRENT FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 09/516,448  
; PRIOR FILING DATE: 2000-02-29  
; NUMBER OF SEQ ID NOS: 2050  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 853  
; LENGTH: 2404  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-833-381-853

Query Match 72.2%; Score 16.6; DB 10; Length 2404;  
Best Local Similarity 82.6%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGCCACAT 23  
||||| | ||||| | |||  
Db 251 ACCAGGAGCCTCGTGGTCCAGAT 229

RESULT 11  
US-09-880-107-3049  
; Sequence 3049, Application US/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3049  
; LENGTH: 339  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:



; OTHER INFORMATION: Genbank Accession No. US20020142981A1 T10108  
US-09-880-107-3049

Query Match 71.3%; Score 16.4; DB 10; Length 339;  
Best Local Similarity 94.4%; Pred. No. 37;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 GCGGCTCTCGTGGCCACA 22  
|||||  
DB 322 GCGGCTCTCGTGGCCAAA 339

## RESULT 12

US-09-764-847-1616/c  
; Sequence 1616, Application US/09764847  
; Patent No. US20020132767A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009  
; CURRENT APPLICATION NUMBER: US/09/764,847  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2003  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1616  
; LENGTH: 10378  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-847-1616

Query Match 70.4%; Score 16.2; DB 10; Length 10378;  
Best Local Similarity 85.7%; Pred. No. 50;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 CAGGAGTCTCTGCGGCCACAT 23  
|||||  
DB 4929 CAGGAGTCTCTGCGGCCACAT 4909

## RESULT 13

US-09-764-847-1615/c  
; Sequence 1615, Application US/09764847  
; Patent No. US20020132767A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009  
; CURRENT APPLICATION NUMBER: US/09/764,847  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2003  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1615  
; LENGTH: 10472  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-847-1615

Query Match 70.4%; Score 16.2; DB 10; Length 10472;  
Best Local Similarity 85.7%; Pred. No. 50;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 CAGGAGTCTCTGCGGCCACAT 23  
|||||  
DB 272 CAGGAGTCTCTGCGGCCACAT 252

## RESULT 14

US-09-764-870-555  
; Sequence 555, Application US/09764870  
; Patent No. US20020042386A1  
; GENERAL INFORMATION:

; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT214  
; CURRENT APPLICATION NUMBER: US/09/764,870  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 646  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 555  
; LENGTH: 10957  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-870-555

Query Match 69.6%; Score 16; DB 10; Length 10957;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTCTCGTGGCCACAT 23  
|||||  
DB 8212 GTCTCGTGGCCACAT 8227

## RESULT 15

US-09-764-853-903  
; Sequence 903, Application US/09764853  
; Patent No. US20020090672A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PJ206  
; CURRENT APPLICATION NUMBER: US/09/764,853  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 939  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 903  
; LENGTH: 10957  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-853-903

Query Match 69.6%; Score 16; DB 10; Length 10957;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTCTCGTGGCCACAT 23  
|||||  
DB 8212 GTCTCGTGGCCACAT 8227

## RESULT 16

US-09-764-870-556/c  
; Sequence 556, Application US/09764870  
; Patent No. US20020042386A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT214  
; CURRENT APPLICATION NUMBER: US/09/764,870  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 646  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 556  
; LENGTH: 13217  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-870-556

Query Match 69.6%; Score 16; DB 10; Length 13217;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY      8 GTCTCGTGGGCCACAT 23
|||||
Db      5005 GTCTCGTGGGCCACAT 4990
|||||

RESULT 17
US-09-764-853-904
; Sequence 904, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 904
; LENGTH: 13217
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-853-904

Query Match      69.6%; Score 16; DB 10; Length 13217;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 GTCTCGTGGGCCACAT 23
|||||
Db      8213 GTCTCGTGGGCCACAT 8228
|||||

RESULT 18
US-09-764-870-554
; Sequence 554, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 554
; LENGTH: 13220
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-870-554

Query Match      69.6%; Score 16; DB 10; Length 13220;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 GTCTCGTGGGCCACAT 23
|||||
Db      8210 GTCTCGTGGGCCACAT 8225
|||||

RESULT 19
US-09-764-853-902
; Sequence 902, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 902
; LENGTH: 13220
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-853-902

Query Match      69.6%; Score 16; DB 10; Length 13220;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 GTCTCGTGGGCCACAT 23
|||||
Db      8210 GTCTCGTGGGCCACAT 8225
|||||

RESULT 20
US-09-864-761-18859/c
; Sequence 18859, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18859
; LENGTH: 305
; TYPE: DNA
; ORGANISM: Homo sapiens
```

FEATURE:  
OTHER INFORMATION: MAP TO AL078604.10  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3  
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 25  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.9  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2  
OTHER INFORMATION: NT HIT: Y13614.1, EVALUE 7.40e-01  
OTHER INFORMATION: SWISSPROT HIT: P50061, EVALUE 1.90e+00  
OTHER INFORMATION: EST\_HUMAN HIT: A1817915.1, EVALUE 1.10e+00  
US-09-864-761-18859

Query Match 67.8%; Score 15.6; DB 10; Length 305;  
Best Local Similarity 81.8%; Pred. No. 88;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CCAGCGCTCTCGTGGCCACAT 23  
||||| ||||| ||||| |||||  
Db 98 CCAGCCTTCTGGTGCACAT 77

RESULT 21  
US-09-815-343-1122  
Sequence 1122, Application US/09815343  
Patent No. US20010055596A1  
GENERAL INFORMATION:  
APPLICANT: Meagher, Madeleine  
APPLICANT: Xu, Jiangchun  
APPLICANT: King, Gordon E.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER  
FILE REFERENCE: 210121.504  
CURRENT APPLICATION NUMBER: US/09/815.343  
CURRENT FILING DATE: 2001-03-22  
NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1122  
LENGTH: 345  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-815-343-1122

Query Match 67.8%; Score 15.6; DB 10; Length 345;  
Best Local Similarity 81.8%; Pred. No. 89;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGCCACA 22  
|| ||||| || ||||| || |||||  
Db 68 ACAGGCGCACTTGTGGCCATA 89

RESULT 22  
US-09-864-761-2116/c  
Sequence 2116, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aecomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864.761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 2116  
LENGTH: 475  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL078604.10  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3  
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 25  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.9  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2  
US-09-864-761-2116

Query Match 67.8%; Score 15.6; DB 10; Length 475;  
Best Local Similarity 81.8%; Pred. No. 89;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CCAGCGCTCTCGTGGCCACAT 23  
||||| ||||| ||||| |||||  
Db 367 CCAGCCTTCTGGTGCACAT 346

RESULT 23  
US-09-864-761-6316  
Sequence 6316, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY



QY 2 CCAGCGCTCTCGTGGGCCACAT 23  
||||||| ||||| |||  
Db 246 CCAGCGTACCGTGGGGGCAT 267

Search completed: December 2, 2002, 22:25:33  
Job time : 79 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2002, 20:57:49 ; Search time 2080 Seconds  
(without alignments)  
179.085 Million cell updates/sec

Title: US-09-700-906A-3

Perfect score: 23

Sequence: 1 accaggcgtctcgtggccacat 23

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32135990

Minimum DB seq length: 66

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpi:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_Other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23	100.0	862	9	AL527673
C 2	22	95.7	248	9	AA355973
C 3	21	91.3	127	9	AA279738
C 4	18.8	81.7	954	17	CNS01TEI
C 5	18.4	80.0	625	10	BE262736
C 6	17.8	77.4	934	14	BM802340

C 7	17.8	77.4	1017	17	CNS03631
C 8	17.4	75.7	424	14	T85624
C 9	17.4	75.7	522	17	AQ589814
C 10	17.4	75.7	528	9	AA118418
C 11	17.4	75.7	1122	17	CNS06R1T
C 12	17.2	74.8	489	10	AV625131
C 13	17.2	74.8	511	10	AV624388
C 14	17.2	74.8	636	10	BB614523
C 15	17.2	74.8	696	17	AZ878371
C 16	17.2	74.8	699	17	AZ427383
C 17	17.2	74.8	724	12	EG476962
C 18	17.2	74.8	788	12	EG475664
C 19	17.2	74.8	921	12	BF101715
C 20	17.2	74.8	929	12	BF983345
C 21	17.2	74.8	1101	12	BE871682
C 22	17.2	74.8	1201	17	CNS002XC
C 23	17	73.9	984	17	CNS04073
C 24	16.8	73.0	341	12	BF844922
C 25	16.8	73.0	379	12	BF844924
C 26	16.8	73.0	478	12	BF751852
C 27	16.8	73.0	480	13	BM049814
C 28	16.8	73.0	488	12	BF512290
C 29	16.8	73.0	520	17	AQ333669
C 30	16.8	73.0	541	10	AV663489
C 31	16.8	73.0	563	17	AQ840043
C 32	16.8	73.0	585	10	BE281235
C 33	16.8	73.0	609	17	AQ288055
C 34	16.8	73.0	635	17	AG166702
C 35	16.8	73.0	822	17	AQ751572
C 36	16.8	73.0	944	13	BT242484
C 37	16.8	73.0	973	12	BG343568
C 38	16.8	73.0	993	12	BG394057
C 39	16.6	72.2	110	12	BF757336
C 40	16.6	72.2	210	9	AV113795
C 41	16.6	72.2	261	10	BB591969
C 42	16.6	72.2	305	10	AV749159
C 43	16.6	72.2	328	9	AA297240
C 44	16.6	72.2	409	10	AW961536
C 45	16.6	72.2	415	9	AA463325

#### ALIGNMENTS

RESULT 1  
AL527673/C  
LOCUS AL527673 862 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL527673 LTI\_NFL003\_NBC3 Homo sapiens cDNA clone CS0DC026YF17 5  
prime, mRNA sequence.  
ACCESSION AL527673  
VERSION AL527673.1 GI:12791166  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
FEATURES  
Location/Qualifiers  
source  
1..862

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DC026YF17"  
/clone\_lib="LTI\_NFL003\_NBC3"  
/sex="male"  
/tissue\_type="neuroblastoma cells"  
/lab\_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 209 a 208 c 251 g 189 t 5 others  
ORIGIN

Query Match 100.0%; Score 23; DB 9; Length 862;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACCAGCGCTCTCGTGGCCACAT 23  
|||||

Db 398 ACCAGCGCTCTCGTGGCCACAT 376

## RESULT 2

AA355973/c 248 bp mRNA linear EST 21-APR-1997  
LOCUS EST64454 Jurkat T-cells VI Homo sapiens cDNA 5' end similar to  
DEFINITION KI-67 antigen, mRNA sequence.

ACCESSION AA355973

VERSION AA355973.1 GI:2008291

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 248)  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## AUTHORS

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult  
,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White  
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,  
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald  
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,  
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,  
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,  
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,  
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
Dinke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,  
,Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,  
Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,  
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon  
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and  
Venter,J.C.

TITLE Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence

JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)

MEDLINE 96026280

COMMENT Other\_ESTs: THC169067

Contact: Kerlavage, AR

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (http://www.tigr.org/tdb/hgi/hgi.html)

Seq primer: M13 Reverse.

## FEATURES

source

1. .248

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):158103"

/db\_xref="taxon:9606"

/clone\_lib="Jurkat T-cells VI"

/cell\_type="T-lymphocyte"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI"

BASE COUNT 44 a 68 c 70 g 60 t 6 others

ORIGIN

Query Match 95.7%; Score 22; DB 9; Length 248;

Best Local Similarity 95.7%; Pred. No. 15;

Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ACCAGCGCTCTCGTGGCCACAT 23

|||||

Db 211 ACCAGCGCTCTCGTGGCCANAT 189

## RESULT 3

AA279738/c

LOCUS AA279738

DEFINITION AA279738

ACCESSION AA279738

VERSION AA279738.1 GI:1921203

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 127)

AUTHORS NCI-CGAP

TITLE http://www.ncbi.nlm.nih.gov/ncicgap.

JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

COMMENT Tumor Gene Index

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Insert Length: 1112 Std Error: 0.00

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 1.

Location/Qualifiers

1. .127

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:704972"

/clone\_lib="NCI-CGAP\_GCBI"

/tissue\_type="germinal center B cell"

/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA

was prepared from human tonsillar cells enriched for

germinal center B cells by flow sorting (CD20+, IgD-),

provided by Dr. Louis M. Staudt (NCI), Dr. David Allman

(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was

primed with a Not I - oligo(dT) primer

[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3'

]. Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7T3 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 27 a 34 c 31 g 35 t

ORIGIN

Query Match 91.3%; Score 21; DB 9; Length 127;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACCAGCGCTCTCGTGGCCAC 21

|||||

Db 21 ACCAGCGCTCTCGTGGCCAC 1







**SOURCE**  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE**  
**AUTHORS** 1 (bases 1 to 522)  
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and  
 Hood,L.  
**TITLE** Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
**MEDLINE** 99380589  
**COMMENT** Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones may be purchased from Research Genetics (info@resgen.com).  
 BAC end Web Server: <http://www.htsc.washington.edu>  
 plate: 2136 row: F column: 16  
 Seq primer: M13 Reverse  
 Class: BAC ends  
 High quality sequence stop: 522.  
**FEATURES**  
**source**  
 1..522  
 /location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate-2136 Col-16 Row-F"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in  
 E-Coli DH10B"  
**BASE COUNT** 147 a 117 c 83 g 175 t  
**ORIGIN**  
 Query Match 75.7%; Score 17.4; DB 17; Length 522;  
 Best Local Similarity 94.7%; Pred. NO. 1.7e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 5 GCGGCTCGTGGGCACAT 23  
 ||||| ||||| ||||| |||||  
 Db 230 GCGGTTTCGTGGGCACAT 212  
**RESULT** 10  
**LOCUS** AA118418  
**DEFINITION** mp54e07.r1 Soares\_thymus\_2NbWT Mus musculus cDNA clone IMAGE:573060  
 5', mRNA sequence.  
**ACCESSION** AA118418  
**VERSION** AA118418.1 GI:1676242  
**KEYWORDS** EST.  
**SOURCE** house mouse.  
**ORGANISM** Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**REFERENCE**  
**AUTHORS** 1 (bases 1 to 528)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
**TITLE** The WashU-HMI Mouse EST Project  
**JOURNAL** Unpublished (1996)  
**COMMENT** Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:347708  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 469.  
**FEATURES**  
**source**  
 1..528  
 /location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:573060"  
 /clone\_lib="Soares\_thymus\_2NbWT"  
 /sex="male"  
 /tissue\_type="Thymus"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer 15'  
 TGTTCACCAATCTGAAGTGGAGCGCGCGTGTGTGTGTGTGTGTGTGTGTGT  
 3'); double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. RNA  
 provided by Dr. Bertrand Jordan. Library went through two  
 rounds of normalization, and was constructed by Bento  
 Soares and M.Fatima Bonaldo."  
**BASE COUNT** 105 a 148 c 125 g 150 t  
**ORIGIN**  
 Query Match 75.7%; Score 17.4; DB 9; Length 528;  
 Best Local Similarity 94.7%; Pred. NO. 1.7e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 4 AGGCGCTCGTGGGCACCA 22  
 ||||| ||||| ||||| |||||  
 Db 390 AGGCGCTCGTGGGCACCA 372  
**RESULT** 11  
**LOCUS** CNS06R1T  
**DEFINITION** T7 end of clone AW0AA010G07 of library AW0AA from strain CLIB 89 of  
 Yarrowia lipolytica, genomic survey sequence.  
**ACCESSION** AL411399  
**VERSION** AL411399.1 GI:12180780  
**KEYWORDS** GSS  
**SOURCE** Yarrowia lipolytica.  
**ORGANISM** Yarrowia lipolytica  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Dipodascaceae; Yarrowia.  
**REFERENCE**  
**AUTHORS** 1 (bases 1 to 1122)  
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,  
 de-Montigny,J., Dujon,B., Durrans,P., Lepingle,A., Liorente,B.,  
 Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
 Wincker,P. and Weissenbach,J.  
**TITLE** Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
**JOURNAL** FEBS Lett. 487 (1), 3-12 (2000)  
**MEDLINE** 20584711  
**COMMENT** PUBMED 11152876  
**REFERENCE** 2 (bases 1 to 1122)  
**AUTHORS** Casaregola,S., Neuveglise,C., Lepingle,A., Bon,E., Feynerol,C.,  
 Artiguenave,F., Wincker,P. and Gaillardin,C.  
**TITLE** Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia  
 lipolytica  
**JOURNAL** FEBS Lett. 487 (1), 95-100 (2000)  
**MEDLINE** 20584727  
**COMMENT** PUBMED 11152892  
**REFERENCE** 3 (bases 1 to 1122)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submission

JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,  
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This GSS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.

## FEATURES

source  
1. 1122  
/organism="Yarrowia lipolytica"  
/strain="CLIB 89"  
/db\_xref="taxon:4952"  
/clone="AW0AA010G07"  
/clone\_lib="AW0AA"  
/note="End : T7"  
misc\_feature  
<484..>783  
/note="similar to Saccharomyces cerevisiae ORF YAL002w [  
VPS8 ; vacuolar sorting protein, 134 kD ]"  
evidence=not\_experimental

BASE COUNT 262 a 308 c 257 g 292 t 3 others  
ORIGIN

Query Match 75.7%; Score 17.4; DB 17; Length 1122;  
Best Local Similarity 85.7%; Pred. No. 2e+03;  
Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGCCAC 21

Db 1102 ATCTGGCTTCGTGGGCCAC 1122

RESULT 12  
AV625131 489 bp mRNA linear EST 15-DEC-2000  
LOCUS AV625131 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas  
reinhardtii cDNA clone LC088c08\_r 5', mRNA sequence.

ACCESSION AV625131  
VERSION AV625131.1 GI:10774308  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii.  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 489)  
AUTHORS Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,  
Nakamura,Y. and Tabata,S.  
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2  
adapted cells of Chlamydomonas reinhardtii  
DNA Res. 7 (5), 305-307 (2000)

COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
Location/Qualifiers  
1. 489

FEATURES source  
1. 489  
/organism="Chlamydomonas reinhardtii"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"  
/note="vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI; The cDNA library was constructed from cells cultured  
in a carbon stress acclimatized condition in which carbon  
dioxide concentration in the bubbling gas was changed from  
5% to 0.04%"

BASE COUNT 90 a 168 c 146 g 85 t  
ORIGIN

Query Match 74.8%; Score 17.2; DB 10; Length 489;  
Best Local Similarity 86.4%; Pred. No. 2e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGGCCACA 22

Db 308 ACCCGCGCTTCGTGGGCCACA 329

## RESULT 13

AV624388 511 bp mRNA linear EST 15-DEC-2000  
LOCUS AV624388 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas  
reinhardtii cDNA clone LC077b01\_r 5', mRNA sequence.

ACCESSION AV624388

VERSION AV624388.1 GI:10773565

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii.

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 511)

AUTHORS Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,

Nakamura,Y. and Tabata,S.

TITLE Generation of expressed sequence tags from low-CO2 and high-CO2

adapted cells of Chlamydomonas reinhardtii

DNA Res. 7 (5), 305-307 (2000)

MEDLINE 20539644

COMMENT Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

## FEATURES

source  
1. 511

/organism="Chlamydomonas reinhardtii"

/strain="C9"

/db\_xref="taxon:3055"

/clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"

/note="vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:

XhoI; The cDNA library was constructed from cells cultured

in a carbon stress acclimatized condition in which carbon

dioxide concentration in the bubbling gas was changed from

5% to 0.04%"

BASE COUNT 94 a 174 c 153 g 90 t

## ORIGIN

Query Match 74.8%; Score 17.2; DB 10; Length 511;  
Best Local Similarity 86.4%; Pred. No. 2e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGGCCACA 22

Db 310 ACCCGCGCTTCGTGGGCCACA 331

## RESULT 14

BB614523 636 bp mRNA linear EST 26-OCT-2001  
LOCUS BB614523 RIKEN full-length enriched, adult male testis Mus musculus  
DEFINITION BB614523 RIKEN full-length enriched, adult male testis Mus musculus  
cDNA clone 4921508F21 5', mRNA sequence.

ACCESSION BB614523

VERSION BB614523.1 GI:16454931

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus.

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 636)





Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LCM1402 row: k column: 04  
High quality sequence stop: 710.  
Location/Qualifiers

#### FEATURES

source

```
1..788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:463927"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into ECORI/XhoI sites using the following 5'
adaptor: GGCAGGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 156 a 211 c 246 g 175 t

ORIGIN

Query Match 74.8%; Score 17.2; DB 12; Length 788;  
Best Local Similarity 86.4%; Pred. No. 2.2e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACCAGCGCTCTGGGCCACA 22

||||||| || |||||||||

Db 726 ACCAGCAACTGTGGGCCACA 705

#### RESULT 19

BF101715/c

LOCUS

DEFINITION 601753449f1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:3980845 5',  
EST.

ACCESSION BF101715

VERSION BF101715.1 GI:10884241

KEYWORDS

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 921)

NIH-MGC <http://mgc.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LAM9177 row: b column: 14

High quality sequence stop: 624.

Location/Qualifiers

1..921

/organism="Mus musculus"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:3980845"

/clone\_lib="NCI\_CGAP\_Mam1"

/tissue\_type="tumor, biopsy sample"

/dev\_stages="10 months, virgin"

/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;

Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigator

providing samples: Gilbert Smith, NIH

247 a 287 c 196 g 191 t

BASE COUNT

ORIGIN

Query Match 74.8%; Score 17.2; DB 12; Length 921;

Best Local Similarity 86.4%; Pred. No. 2.3e+03;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCAGCGCTCTGTGGGCCACAT 23

||||| |||||||||

Db 664 CCGTCATCTGTGGGCCACAT 643

#### RESULT 20

BF983345/c

LOCUS

DEFINITION 602305223f1 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:4396612 5',  
EST.

ACCESSION BF983345

VERSION BF983345.1 GI:12386157

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 929)

NIH-MGC <http://mgc.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LAM10095 row: n column: 05

High quality sequence stop: 632.

Location/Qualifiers

1..929

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4396612"

/clone\_lib="NIH\_MGC\_88"

/tissue\_type="duodenal adenocarcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: small intestine; Vector: pCMV-SPORT6;

Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally;

oligo-dT primed. Average insert size 1.767 kb. Library

enriched for full-length clones and constructed by Life

Technologies. Note: this is a NIH-MGC Library."

229 a 247 c 283 g 170 t

BASE COUNT

ORIGIN

Query Match 74.8%; Score 17.2; DB 12; Length 929;

Best Local Similarity 86.4%; Pred. No. 2.3e+03;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACCAGCGCTCTGTGGGCCACA 22

||||| |||||||||

Db 730 ATCCGCGTCTCATGGGCCACA 709

#### RESULT 21

BE871682

LOCUS

DEFINITION 601449542f1 NIH\_MGC\_65 Homo sapiens cDNA clone IMAGE:3853379 5',  
EST.

ACCESSION BE871682

VERSION BE871682.1 GI:10320458

KEYWORDS

EST.

SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 1101)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMLL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMLL at:  
<http://image.llnl.gov>  
Plate: LLM9577 row: c column: 12  
High quality sequence stop: 630.

FEATURES  
source  
Location/Qualifiers  
1..1101  
    /organism="Homo sapiens"  
    /db\_xref="taxon:9606"  
    /clone IMAGE:3853379  
    /clone\_lib="NIH\_MGC\_65"  
    /tissue\_type="adenocarcinoma"  
    /lab\_host="DH10B (phage-resistant)"  
    /note="Organ: colon; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.8 kb. Library constructed by Life  
Technologies."

BASE COUNT  
ORIGIN  
245 a 395 c 283 g 178 t

Query Match 74.8%; Score 17.2; DB 12; Length 1101;  
Best Local Similarity 86.4%; Pred. No. 2.4e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACCAGGGCTCTCGGGGCCACA 22  
|| ||||| ||||||||| ||  
Db 384 ACGAGGCGACTCGGGGCCATA 405

RESULT 22  
CNS002XC/c  
LOCUS  
CNS002XC 1201 bp DNA linear GSS 26-JUL-1999  
Drosophila melanogaster genome survey sequence Sf6 end of BAC  
BACN03K17 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
AL098250.1 GI:5609861  
GSS.  
Drosophila melanogaster.  
Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
1 (bases 1 to 1201)  
Genoscope.  
Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ;  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
<http://www.edgp.ebi.ac.uk> . This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
peloBAC11.

FEATURES  
source  
Location/Qualifiers  
1..1201



```
RESULT 24
BF844922
LOCUS          BF844922          341 bp      mRNA          linear          EST 13-JAN-2001
DEFINITION    RC4-HT0978-221200-021-h04 HT0978 Homo sapiens cDNA, mRNA sequence.
ACCESSION     BF844922
VERSION       BF844922.1  GI:12200923
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS       1 (bases 1 to 341)
               Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
               Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
               Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
               Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
               M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
               Simpson,A.J.
               Shotgun sequencing of the human transcriptome with ORF expressed
               sequence tags
               Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT       Contact: Simpson A.J.G.
               Laboratory of Cancer Genetics
               Ludwig Institute for Cancer Research
               Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
               Brazil
               Tel: +55-11-2704922
               Fax: +55-11-2707001
               Email: asimpson@ludwig.org.br
               This sequence was derived from the FAPESP/LICR Human Cancer Genome
               Project. This entry can be seen in the following URL
               (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC4&t2=RC4-HT0978-
               221200-021-h04&t3=2000-12-22&t4=1)
               Seq primer: puc 18 forward
               High quality sequence start: 20
               High quality sequence stop: 340.
               Location/Qualifiers
                   1..341
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone_lib="HT0978"
                   /dev_stage="Adult"
                   /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                   Site_2: SmaI; A mini-library was made by cloning products
                   derived from ORESTES PCR (U.S. Letters Patent application
                   No. 196,716 - Ludwig Institute for Cancer Research)
                   profiles into the puc 18 vector. Reverse transcription of
                   tissue mRNA and cDNA amplification were performed under
                   low stringency conditions."
                   63 a 105 c 100 g 73 t

BASE COUNT    63 a 105 c 100 g 73 t
ORIGIN
Query Match   73.0%; Score 16.8; DB 12; Length 341;
Best Local Similarity 90.0%; Pred. No. 2.8e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGCCA 20
||||| ||||| ||||| |||||
Db 96 ACCAGGTGTCTCTCTGGCCA 115

RESULT 25
BF844924
LOCUS          BF844924          379 bp      mRNA          linear          EST 13-JAN-2001
DEFINITION    RC4-HT0978-221200-021-h08 HT0978 Homo sapiens cDNA, mRNA sequence.
ACCESSION     BF844924
VERSION       BF844924.1  GI:12200929
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS       1 (bases 1 to 379)
               Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
               Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
               Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
               Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
               M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
               Simpson,A.J.
               Shotgun sequencing of the human transcriptome with ORF expressed
               sequence tags
               Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT       Contact: Simpson A.J.G.
               Laboratory of Cancer Genetics
               Ludwig Institute for Cancer Research
               Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
               Brazil
               Tel: +55-11-2704922
               Fax: +55-11-2707001
               Email: asimpson@ludwig.org.br
               This sequence was derived from the FAPESP/LICR Human Cancer Genome
               Project. This entry can be seen in the following URL
               (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC4&t2=RC4-HT0978-
               221200-021-h04&t3=2000-12-22&t4=1)
               Seq primer: puc 18 forward
               High quality sequence start: 20
               High quality sequence stop: 379.
               Location/Qualifiers
                   1..379
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone_lib="HT0978"
                   /dev_stage="Adult"
                   /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                   Site_2: SmaI; A mini-library was made by cloning products
                   derived from ORESTES PCR (U.S. Letters Patent application
                   No. 196,716 - Ludwig Institute for Cancer Research)
                   profiles into the puc 18 vector. Reverse transcription of
                   tissue mRNA and cDNA amplification were performed under
                   low stringency conditions."
                   73 a 115 c 114 g 77 t

BASE COUNT    73 a 115 c 114 g 77 t
ORIGIN
Query Match   73.0%; Score 16.8; DB 12; Length 379;
Best Local Similarity 90.0%; Pred. No. 2.8e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ACCAGGCGTCTCGTGGCCA 20
||||| ||||| ||||| |||||
Db 143 ACCAGGTGTCTCTCTGGCCA 162

Search completed: December 2, 2002, 22:24:07
Job time : 2095 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2002, 22:47:00 ; Search time 15857 Seconds  
(without alignments)  
17923.819 Million cell updates/sec

Title: US-09-700-906A-1\_COPY\_197\_9962  
Perfect score: 9766  
Sequence: 1 atgtggccacgacgcct.....tcataggacagtgaagata 9766

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 918046

Minimum DB seq length: 0  
Maximum DB seq length: 66

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htg\_mus.\*  
34: em.htg\_pln.\*  
35: em.htg\_rod.\*  
36: em.htg\_mam.\*  
37: em.htg\_vrt.\*  
38: em\_sy.\*  
39: em.htgo\_hum.\*  
40: em.htgo\_mus.\*  
41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	51	0.5	51	6	AX204486	AX204486 Sequence
c 2	26.2	0.3	51	6	AX204495	AX204495 Sequence
c 3	26	0.3	65	6	AX483033	AX483033 Sequence
c 4	25.2	0.3	66	9	AF189407	AF189407 Homo sapi
5	25	0.3	50	6	AR032788	AR032788 Sequence
6	25	0.3	50	6	AR209452	AR209452 Sequence
7	25	0.3	50	6	I29528	I29528 Sequence 40
8	25	0.3	50	6	I91202	I91202 Sequence 40
9	25	0.3	50	6	CMV17101	Y17101 Cladobotryu
c 10	24.8	0.3	50	6	I42230	I42230 Sequence 43
11	24.8	0.3	51	6	AX118213	AX118213 Sequence
12	24.8	0.3	66	6	AR209041	AR209041 Sequence
13	24.6	0.3	60	6	E22184	E22184 Yeast capab
14	24.4	0.2	63	8	CMV17098	Y17098 Cladobotryu
c 15	24.2	0.2	60	9	HSEZRNG18	AF053448 Homo sapi
16	24.2	0.2	62	6	I31144	I31144 Sequence 56
17	24	0.2	51	9	HSU14016	U14016 Human T cel
18	24	0.2	54	6	AR041096	AR041096 Sequence
19	24	0.2	54	6	AR060704	AR060704 Sequence
20	24	0.2	54	6	AR063137	AR063137 Sequence
21	24	0.2	54	6	AR117026	AR117026 Sequence
22	24	0.2	54	6	I64788	I64788 Sequence 13
c 23	24	0.2	58	6	AR098671	AR098671 Sequence
c 24	24	0.2	58	6	AR204745	AR204745 Sequence
25	23.6	0.2	58	9	HSU96955	U96955 Homo sapien
26	23.6	0.2	58	9	HSU96956	U96956 Homo sapien
27	23.6	0.2	60	9	HSU96962	U96962 Homo sapien
28	23.6	0.2	62	9	HSU96958	U96958 Homo sapien
29	23.6	0.2	64	9	HSU96959	U96959 Homo sapien
30	23.6	0.2	66	9	HSU96960	U96960 Homo sapien
31	23.4	0.2	50	6	I31257	I31257 Sequence 16
32	23.4	0.2	58	6	AX133328	AX133328 Sequence
33	23.4	0.2	61	6	AX270718	AX270718 Sequence
34	23.4	0.2	61	6	AX272249	AX272249 Sequence
35	23.4	0.2	61	6	I31370	I31370 Sequence 28
36	23.4	0.2	62	9	S52152S31	S52228 CD11b-leuko
c 37	23.4	0.2	65	6	AX483325	AX483325 Sequence
c 38	23.4	0.2	65	6	AX483936	AX483936 Sequence
c 39	23.2	0.2	46	6	AR081408	AR081408 Sequence
c 40	23.2	0.2	49	6	AX026790	AX026790 Sequence
c 41	23.2	0.2	51	6	AX159132	AX159132 Sequence
c 42	23.2	0.2	56	6	I06069	I06069 Sequence 16
c 43	23.2	0.2	57	6	AR135505	AR135505 Sequence
c 44	23.2	0.2	64	6	AX381699	AX381699 Sequence
c 45	23	0.2	23	6	AX009578	AX009578 Sequence

ALIGNMENTS

RESULT 1  
AX204486  
LOCUS AX204486 51 bp DNA linear PAT 30-AUG-2001  
DEFINITION Sequence 592 from Patent WO0148245.  
ACCESSION AX204486  
VERSION AX204486.1 GI:15394046  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 51)  
AUTHORS Shimkets, R.A. and Leach, M.  
TITLE Nucleic acids containing single nucleotide polymorphisms and methods of use thereof



<p>RESULT 5</p> <p>LOCUS AR032788</p> <p>DEFINITION Sequence 400 from patent US 5869241.</p> <p>ACCESSION AR032788</p> <p>VERSION AR032788.1 GI:5948393</p> <p>KEYWORDS</p> <p>SOURCE Unknown.</p> <p>ORGANISM Unknown.</p> <p>REFERENCE 1 (bases 1 to 50)</p> <p>AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.</p> <p>TITLE Method of determining DNA sequence preference of a DNA-binding molecule</p> <p>JOURNAL Patent: US 5869241-A 400 09-FEB-1999;</p> <p>FEATURES</p> <p>source Location/Qualifiers</p> <p>1..50</p> <p>/organism="unknown"</p> <p>BASE COUNT 20 a 12 c 14 g 4 t</p> <p>ORIGIN</p> <p>Query Match 0.3%; Score 25; DB 6; Length 50;</p> <p>Best Local Similarity 75.6%; Pred. No. 9.2e+05;</p> <p>Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;</p> <p>QY 6062 CGTCAGGGAAGACCCACACAGACACAGAGGCGCAGCA 6102</p> <p>Db 9 CGCAGTGAAGATATAAAGGCCACACAGAGGCGCAGCA 49</p>	<p>RESULT 6</p> <p>LOCUS AR209452</p> <p>DEFINITION Sequence 400 from patent US 6384208.</p> <p>ACCESSION AR209452</p> <p>VERSION AR209452.1 GI:21510880</p> <p>KEYWORDS</p> <p>SOURCE Unknown.</p> <p>ORGANISM Unknown.</p> <p>REFERENCE 1 (bases 1 to 50)</p> <p>AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.</p> <p>TITLE Sequence directed DNA binding molecules compositions and methods</p> <p>JOURNAL Patent: US 6384208-A 400 07-MAY-2002;</p> <p>FEATURES</p> <p>source Location/Qualifiers</p> <p>1..50</p> <p>/organism="unknown"</p> <p>BASE COUNT 20 a 12 c 14 g 4 t</p> <p>ORIGIN</p> <p>Query Match 0.3%; Score 25; DB 6; Length 50;</p> <p>Best Local Similarity 75.6%; Pred. No. 9.2e+05;</p> <p>Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;</p> <p>QY 6062 CGTCAGGGAAGACCCACACAGACACAGAGGCGCAGCA 6102</p> <p>Db 9 CGCAGTGAAGATATAAAGGCCACACAGAGGCGCAGCA 49</p>	<p>RESULT 7</p> <p>LOCUS I29528</p> <p>DEFINITION Sequence 400 from patent US 5578444.</p> <p>ACCESSION I29528</p> <p>VERSION I29528.1 GI:1820319</p> <p>KEYWORDS</p> <p>SOURCE Unknown.</p> <p>ORGANISM Unknown.</p> <p>REFERENCE 1 (bases 1 to 50)</p> <p>AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.</p>	<p>RESULT 8</p> <p>LOCUS I91202</p> <p>DEFINITION Sequence 400 from patent US 5726014.</p> <p>ACCESSION I91202</p> <p>VERSION I91202.1 GI:3935672</p> <p>KEYWORDS</p> <p>SOURCE Unknown.</p> <p>ORGANISM Unknown.</p> <p>REFERENCE 1 (bases 1 to 50)</p> <p>AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M. and Turin,L.M.</p> <p>TITLE Screening assay for the detection of DNA-binding molecules</p> <p>JOURNAL Patent: US 5726014-A 400 10-MAR-1998;</p> <p>FEATURES</p> <p>source Location/Qualifiers</p> <p>1..50</p> <p>/organism="unknown"</p> <p>BASE COUNT 20 a 12 c 14 g 4 t</p> <p>ORIGIN</p> <p>Query Match 0.3%; Score 25; DB 6; Length 50;</p> <p>Best Local Similarity 75.6%; Pred. No. 9.2e+05;</p> <p>Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;</p> <p>QY 6062 CGTCAGGGAAGACCCACACAGACACAGAGGCGCAGCA 6102</p> <p>Db 9 CGCAGTGAAGATATAAAGGCCACACAGAGGCGCAGCA 49</p>	<p>RESULT 9</p> <p>LOCUS CMY17101</p> <p>DEFINITION Cladobotryum mycophilum microsatellite DNA, isolate IMI 26134.</p> <p>ACCESSION Y17101</p> <p>VERSION Y17101.1 GI:3688265</p> <p>KEYWORDS microsatellite.</p> <p>SOURCE Hypomyces odoratus</p> <p>ORGANISM Hypomyces odoratus</p> <p>REFERENCE 1 (bases 1 to 61)</p> <p>AUTHORS McKay,G.J., Egan,D., Morris,E., Scott,C. and Brown,A.E.</p> <p>TITLE Genetic and morphological characterization of Cladobotryum species causing cobweb disease of mushrooms</p> <p>JOURNAL Appl. Environ. Microbiol. 65 (2), 606-610 (1999)</p> <p>MEDLINE 99124595</p> <p>PUBMED 9925589</p> <p>REFERENCE 2 (bases 1 to 61)</p> <p>AUTHORS McKay,G.J.</p> <p>TITLE Direct Submission</p> <p>JOURNAL Submitted (01-APR-1998) G.J. McKay, Queens University Belfast, Department of Applied Plant Science, Newforge Lane, Belfast, N. Ireland B9 5PX, UK</p> <p>FEATURES</p> <p>source Location/Qualifiers</p> <p>1..61</p> <p>/organism="Hypomyces odoratus"</p>	<p>RESULT 10</p> <p>LOCUS CMY17101</p> <p>DEFINITION Cladobotryum mycophilum microsatellite DNA, isolate IMI 26134.</p> <p>ACCESSION Y17101</p> <p>VERSION Y17101.1 GI:3688265</p> <p>KEYWORDS microsatellite.</p> <p>SOURCE Hypomyces odoratus</p> <p>ORGANISM Hypomyces odoratus</p> <p>REFERENCE 1 (bases 1 to 61)</p> <p>AUTHORS McKay,G.J., Egan,D., Morris,E., Scott,C. and Brown,A.E.</p> <p>TITLE Genetic and morphological characterization of Cladobotryum species causing cobweb disease of mushrooms</p> <p>JOURNAL Appl. Environ. Microbiol. 65 (2), 606-610 (1999)</p> <p>MEDLINE 99124595</p> <p>PUBMED 9925589</p> <p>REFERENCE 2 (bases 1 to 61)</p> <p>AUTHORS McKay,G.J.</p> <p>TITLE Direct Submission</p> <p>JOURNAL Submitted (01-APR-1998) G.J. McKay, Queens University Belfast, Department of Applied Plant Science, Newforge Lane, Belfast, N. Ireland B9 5PX, UK</p> <p>FEATURES</p> <p>source Location/Qualifiers</p> <p>1..61</p> <p>/organism="Hypomyces odoratus"</p>
---	---	--	--	---	--

[illegible]



```
source
1..60
/organism="unidentified"
/db_xref="taxon:32644"
23 a 13 c 16 g 8 t

Query Match 0.3%; Score 24.6; DB 6; Length 60;
Best Local Similarity 70.2%; Pred. No. 1.2e+06;
Matches 33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 7957 AAACCAACCCAGTAGAAGAGAACCCAGCAGGAGAGAGCCCAAGAGC 8003
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 7 ATGCAAACTAGTAGACGAGGAAGTTACCGAGACACAGAGAAACGC 53

RESULT 14
CMV17098
LOCUS CMV17098 63 bp DNA linear PLN 24-FEB-1999
DEFINITION Cladobotryum mycophilum microsatellite DNA, isolate CBS 111.92.
ACCESSION Y17098
VERSION Y17098.1 GI:3688262
KEYWORDS microsatellite.
SOURCE Hypomyces odoratus.
ORGANISM Hypomyces odoratus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; Hypocreaceae; Hypomyces.
1 (bases 1 to 63)
McKay,G.J., Egan,D., Morris,E., Scott,C. and Brown,A.E.
Genetic and morphological characterization of Cladobotryum species
causing cobweb disease of mushrooms
Appl. Environ. Microbiol. 65 (2), 606-610 (1999)
JOURNAL
MEDLINE 99124595
PUBMED 9925589
REFERENCE 2 (bases 1 to 63)
McKay,G.J.
AUTHORS McKay,G.J.
DIRECT SUBMISSION
TITLE Submitted (01-APR-1998) G.J. McKay, Queens University Belfast,
Department of Applied Plant Science, Newforge Lane, Belfast, N.
Ireland BT9 5PX, UK
FEATURES
Location/Qualifiers
1..63
/organism="Hypomyces odoratus"
/isolate="CBS 111.92"
/specific_host="Tricholoma terreum"
/db_xref="taxon:76883"
/tissue_type="mycelium"
1..63
satellite
1..63
/note="microsatellite"
29 a 3 c 24 g 7 t

BASE COUNT 29 a 3 c 24 g 7 t
ORIGIN

Query Match 0.2%; Score 24.4; DB 8; Length 63;
Best Local Similarity 63.8%; Pred. No. 1.3e+06;
Matches 37; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 6055 ACACAGACGCTCAGGGAAGACACACAGACACACAGAGAGACGAGATGGAAGA 6112
I IIII I IIII I IIII I IIII I IIII I IIII I IIII
Db 2 ATAAAGCCACATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 59

RESULT 15
HSEZRING18/c
LOCUS HSEZRING18 60 bp DNA linear PRI 06-APR-1999
DEFINITION Homo sapiens ezrin gene, 3' acceptor end, partial intron 9.
ACCESSION AF053448
VERSION AF053448.1 GI:4104505
KEYWORDS 18 of 22
SEGMENT Homo sapiens
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 60)
AUTHORS Majander-Nordenswan,P., Sainio,M., Turunen,O., Jaaskelainen,J.,
Carpen,O., Kere,J. and Vaheiri,A.
TITLE Genomic structure of the human ezrin gene
JOURNAL Hum. Genet. 103 (6), 662-665 (1998)
MEDLINE 99118873
PUBMED 9921900
REFERENCE 2 (bases 1 to 60)
Majander-Nordenswan,P., Sainio,M., Turunen,O., Jaaskelainen,J.,
Carpen,O., Kere,J. and Vaheiri,A.
DIRECT SUBMISSION
TITLE Submitted (01-DEC-1997) Haartman Institute, University of Helsinki,
Haartmaninkatu 3 PL 21, Helsinki FIN-00014, Finland
FEATURES
Location/Qualifiers
1..60
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q25-q26; between 6D6S442 and D6S281"
/clone="P1 genomic clones #6096 and #6097 (Genome Systems,
Inc.)"
<1..60
/note="ezrin; 3' acceptor"
/number=9
6 a 13 c 22 g 19 t

BASE COUNT 6 a 13 c 22 g 19 t
ORIGIN

Query Match 0.2%; Score 24.2; DB 9; Length 60;
Best Local Similarity 78.4%; Pred. No. 1.4e+06;
Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 7389 CTGTAATCTCCACAGCCAGAGTCATTTCAAAACCTCA 7425
III III III IIIII IIIII IIIII IIIII IIIII
Db 60 CTGCAAGAGACAAAGCCAGAGCCATTTCAAAACCTCA 24

Search completed: December 3, 2002, 06:48:19
Job time : 15894 secs
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2002, 22:25:40 ; Search time 1179 Seconds  
(without alignments)  
18653.957 Million cell updates/sec

Title: US-09-700-906A-1\_COPY\_197\_9962  
Perfect score: 9766  
Sequence: 1 atgtggccacgagcgcct.....tcattagggacagtgaagata 9766

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 2313090

Minimum DB seq length: 0  
Maximum DB seq length: 66

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :				N_Geneseq_101002:*				
				1:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*			
				2:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*			
				3:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*			
				4:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*			
				5:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*			
				6:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*			
				7:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*			
				8:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*			
				9:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*			
				10:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*			
				11:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*			
				12:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*			
				13:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*			
				14:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*			
				15:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*			
				16:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*			
				17:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*			
				18:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*			
				19:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*			
				20:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*			
				21:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*			
				22:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*			
				23:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*			
				24:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*			

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description	
1	51	0.5	51	22	AAL27414	Human SNP oligonuc
2	51	0.5	51	22	AAL32696	Human SNP oligonuc
3	51	0.5	51	22	AAL32697	Human SNP oligonuc
4	51	0.5	51	22	AAL32698	Human SNP oligonuc
5	51	0.5	51	22	AAL33289	Human SNP oligonuc
6	51	0.5	51	22	AAL33290	Human SNP oligonuc
7	51	0.5	51	22	AAH79977	Human DNA containi
8	49.4	0.5	51	22	AAL27415	Human SNP oligonuc
9	49.4	0.5	51	22	AAL33291	Human SNP oligonuc

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

10	49.4	0.5	51	22	AAL34120	Human SNP oligonuc
11	34.2	0.4	65	24	ABN51074	Mouse spliced tran
c 12	30.6	0.3	31	19	AAV67602	Nucleotide fragmen
13	28.2	0.3	58	22	AAI64443	SSR motif #3, uni
14	27.8	0.3	53	21	AAA64666	PCR primer used to
c 15	27.4	0.3	36	21	AAA64667	PCR primer used to
16	26.8	0.3	60	24	ABN37508	Human spliced tran
c 17	26.2	0.3	51	22	AAH79986	Human DNA containi
c 18	26.2	0.3	62	21	AAC12223	Human secreted pro
c 19	25.6	0.3	65	21	AAC14733	Human secreted pro
c 20	25.4	0.3	65	24	ABN51781	Mouse spliced tran
21	25	0.3	50	15	AAQ69650	Orangutan epsilon-
22	25	0.3	50	18	AAT64112	Orangutan epsilon-
23	25	0.3	50	20	AAH17400	Test sequence from
24	25	0.3	50	24	ABK82891	DNA binding molecu
25	25	0.3	65	24	ABN27757	Rat spliced transcr
c 26	24.8	0.3	50	11	AAQ06733	:HBV.XT1.61 captur
c 27	24.8	0.3	50	12	AAQ12932	Analyte capture pr
c 28	24.8	0.3	50	18	AAT74380	Analyte capture pr
29	24.8	0.3	51	22	AAH40540	Human SNP flanking
30	24.6	0.3	60	20	AAH15719	PCR primer used to
31	24.6	0.3	65	24	ABN51514	Mouse spliced tran
32	24.2	0.2	51	18	AAV76224	Staphylococcus aur
33	24.2	0.2	62	18	AAT65704	Repeat sequence fr
c 35	24	0.2	24	19	AAH09923	Human biallelic po
c 36	24	0.2	24	19	AAH09349	Human biallelic po
c 37	24	0.2	50	22	AAL34538	Human SNP oligonuc
38	24	0.2	54	16	AAQ87143	pBR322 ori upstrea
39	24	0.2	54	17	AAT32724	Oligo TG62 for cre
40	24	0.2	54	18	AAT60985	Upstream primer fo
41	24	0.2	54	18	AAT73264	Upstream primer TG
42	24	0.2	54	19	AAV58345	Primer TG62 for ha
c 43	24	0.2	54	22	AAC81484	pBR322 ori upstrea
c 44	24	0.2	58	19	AAV59234	Circular template
c 45	23.8	0.2	58	19	AAV12920	Oligonucleotide SE
			47	21	AAZ69060	Human map-related

ALIGNMENTS

RESULT 1  
AAL27414  
ID AAL27414 standard; DNA; 51 BP.  
XX  
AC AAL27414;  
XX  
XX 24-JAN-2002 (first entry)  
DE Human SNP oligonucleotide #622.  
XX

XX Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;  
KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;  
KW amyloid protein; angiotensin; apoptosis related protein; cadherin;  
KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;  
KW complement related protein; cytochrome; kinesin; cytokine; interferon;  
KW interleukin; G-protein coupled receptor; thioesterase; inflammation;  
KW multifactorial disease; autoimmune disease; infection;  
KW nervous system disease; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200147944-A2.  
XX  
XX 05-JUL-2001.  
XX  
XX 28-DEC-2000; 2000WO-US35498.  
XX  
XX 28-DEC-1999; 99US-0173419.  
XX  
XX 27-DEC-2000; 2000US-0173419.  
XX  
XX (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach M;  
 XX WPI; 2001-465210/50.  
 DR Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,  
 XX oncogenes and histones, useful for diagnosing and treating, e.g.  
 PT cancer, autoimmune diseases and infections -  
 PT Claim 1; Page 1565; 4143pp; English.  
 XX  
 XX The present invention relates to oligonucleotides encoding polymorphic  
 CC variants of proteins related to amylases, amyloid proteins, angiotensin,  
 CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,  
 CC histones, kinases, colony stimulating factors, complement related  
 CC proteins, cytochromes, cytokines, interferons, interleukins,  
 CC G-protein coupled receptors and thioesterases. The present sequence is  
 CC one such oligonucleotide. The oligonucleotides and the peptides encoded  
 CC by them may be used in the prevention, diagnosis and treatment of  
 CC diseases associated with inappropriate expression of the proteins listed  
 CC above. Disorders that may be prevented, diagnosed and/or treated include  
 CC multifactorial diseases with a genetic component, such as autoimmune  
 CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,  
 CC systemic lupus erythematosus and Grave's disease), inflammation, cancer  
 CC (e.g. cancers of the bladder, brain, breast, colon and kidney,  
 CC leukaemia), diseases of the nervous system and an infection of pathogenic  
 CC organisms.  
 XX  
 SQ Sequence 51 BP; 20 A; 16 C; 8 G; 7 T; 0 other;  
 Query Match 0.5%; Score 51; DB 22; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 0.0068;  
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6614 TATGCACTGACAGCCGACGCTCAGCAGAAACTACCAAAATAGCTGCA 6664  
 Db 1 TATGCACTGACAGCCGACGCTCAGCAGAAACTACCAAAATAGCTGCA 51  
 RESULT 2  
 AAL32696  
 ID AAL32696 standard; DNA; 51 BP.  
 XX  
 XX AAL32696;  
 AC  
 XX 24-JAN-2002 (first entry)  
 DT  
 XX Human SNP oligonucleotide #5904.  
 DE  
 XX Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;  
 KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;  
 KW amyloid protein; angiotensin; apoptosis related protein; cadherin;  
 KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;  
 KW complement related protein; cytochrome; cytokine; interferon;  
 KW interleukin; G-protein coupled receptor; thioesterase; inflammation;  
 KW multifactorial disease; autoimmune disease; infection;  
 KW nervous system disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200147944-A2.  
 PN  
 XX 05-JUL-2001.  
 PD  
 XX 28-DEC-2000; 2000WO-US35498.  
 XX  
 XX 28-DEC-1999; 99US-0173419.  
 PR 27-DEC-2000; 2000US-0173419.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX Shimkets RA, Leach M;  
 PI WPI; 2001-465210/50.  
 XX  
 DR Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,  
 XX oncogenes and histones, useful for diagnosing and treating, e.g.  
 PT cancer, autoimmune diseases and infections -

XX Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,  
 PT oncogenes and histones, useful for diagnosing and treating, e.g.  
 PT cancer, autoimmune diseases and infections -  
 XX Claim 1; Page 3083; 4143pp; English.  
 PS  
 XX The present invention relates to oligonucleotides encoding polymorphic  
 CC variants of proteins related to amylases, amyloid proteins, angiotensin,  
 CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,  
 CC histones, kinases, colony stimulating factors, complement related  
 CC proteins, cytochromes, cytokines, interferons, interleukins,  
 CC G-protein coupled receptors and thioesterases. The present sequence is  
 CC one such oligonucleotide. The oligonucleotides and the peptides encoded  
 CC by them may be used in the prevention, diagnosis and treatment of  
 CC diseases associated with inappropriate expression of the proteins listed  
 CC above. Disorders that may be prevented, diagnosed and/or treated include  
 CC multifactorial diseases with a genetic component, such as autoimmune  
 CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,  
 CC systemic lupus erythematosus and Grave's disease), inflammation, cancer  
 CC (e.g. cancers of the bladder, brain, breast, colon and kidney,  
 CC leukaemia), diseases of the nervous system and an infection of pathogenic  
 CC organisms.  
 XX  
 SQ Sequence 51 BP; 16 A; 17 C; 9 G; 9 T; 0 other;  
 Query Match 0.5%; Score 51; DB 22; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 0.0068;  
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6588 CTTGAAGAGCTCTTCCAGACACCACTATGCAGTGAAGCCGACGACTCA 6638  
 Db 1 CTTGAAGAGCTCTTCCAGACACCACTATGCAGTGAAGCCGACGACTCA 51  
 RESULT 3  
 AAL32697  
 ID AAL32697 standard; DNA; 51 BP.  
 XX  
 XX AAL32697;  
 AC  
 XX 24-JAN-2002 (first entry)  
 DT  
 XX Human SNP oligonucleotide #5905.  
 DE  
 XX Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;  
 KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;  
 KW amyloid protein; angiotensin; apoptosis related protein; cadherin;  
 KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;  
 KW complement related protein; cytochrome; cytokine; interferon;  
 KW interleukin; G-protein coupled receptor; thioesterase; inflammation;  
 KW multifactorial disease; autoimmune disease; infection;  
 KW nervous system disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200147944-A2.  
 PN  
 XX 05-JUL-2001.  
 PD  
 XX 28-DEC-2000; 2000WO-US35498.  
 XX  
 XX 28-DEC-1999; 99US-0173419.  
 PR 27-DEC-2000; 2000US-0173419.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX Shimkets RA, Leach M;  
 PI WPI; 2001-465210/50.  
 XX  
 DR Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,  
 XX oncogenes and histones, useful for diagnosing and treating, e.g.  
 PT cancer, autoimmune diseases and infections -

PT cancer, autoimmune diseases and infections -  
XX Claim 1; Page 3083; 4143pp; English.  
XX  
XX The present invention relates to oligonucleotides encoding polymorphic  
CC variants of proteins related to amylases, amyloid proteins, angiotensin,  
CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,  
CC histones, kinases, colony stimulating factors, complement related  
CC proteins, cytochromes, cytokines, interferons, interleukins,  
CC G-protein coupled receptors and thioesterases. The present sequence is  
CC one such oligonucleotide. The oligonucleotides and the peptides encoded  
CC by them may be used in the prevention, diagnosis and treatment of  
CC diseases associated with inappropriate expression of the proteins listed  
CC above. Disorders that may be prevented, diagnosed and/or treated include  
CC multifactorial diseases with a genetic component, such as autoimmune  
CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,  
CC systemic lupus erythematosus and Grave's disease), inflammation, cancer  
CC (e.g. cancers of the bladder, brain, breast, colon and kidney,  
CC leukaemia), diseases of the nervous system and an infection of pathogenic  
CC organisms.  
XX  
SQ Sequence 51 BP; 19 A; 14 C; 11 G; 7 T; 0 other;  
Query Match 0.5%; Score 51; DB 22; Length 51;  
Best Local Similarity 100.0%; Pred. No. 0.0068;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9264 AATATCCCTGCGCTCCAGACGCAAGATGAAGACTGAGGCGAGACAGCAAT 9314  
|||||  
Db 1 AATATCCCTGCGCTCCAGACGCAAGATGAAGACTGAGGCGAGACAGCAAT 51  
RESULT 4  
AAL32698  
ID AAL32698 standard; DNA; 51 BP.  
XX  
XX AAL32698;  
AC  
XX  
XX 24-JAN-2002 (first entry)  
DE Human SNP oligonucleotide #5906.  
XX  
XX Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;  
KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;  
KW amyloid protein; angiotensin; apoptosis related protein; cadherin;  
KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;  
KW complement related protein; cytochrome; kinesin; cytokine; interferon;  
KW interleukin; G-protein coupled receptor; thioesterase; inflammation;  
KW multifactorial disease; autoimmune disease; infection;  
KW nervous system disease; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200147944-A2.  
XX  
XX 05-JUL-2001.  
XX  
XX 28-DEC-2000; 2000WO-US35498.  
XX  
XX 28-DEC-1999; 99US-0173419.  
XX 27-DEC-2000; 2000US-0173419.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Shimkets RA, Leach M;  
XX  
XX WPI; 2001-465210/50.  
XX  
XX Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,  
PT oncogenes and histones, useful for diagnosing and treating, e.g.  
PT cancer, autoimmune diseases and infections -  
XX Claim 1; Page 3084; 4143pp; English.  
PS

XX The present invention relates to oligonucleotides encoding polymorphic  
CC variants of proteins related to amylases, amyloid proteins, angiotensin,  
CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,  
CC histones, kinases, colony stimulating factors, complement related  
CC proteins, cytochromes, cytokines, interferons, interleukins,  
CC G-protein coupled receptors and thioesterases. The present sequence is  
CC one such oligonucleotide. The oligonucleotides and the peptides encoded  
CC by them may be used in the prevention, diagnosis and treatment of  
CC diseases associated with inappropriate expression of the proteins listed  
CC above. Disorders that may be prevented, diagnosed and/or treated include  
CC multifactorial diseases with a genetic component, such as autoimmune  
CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,  
CC systemic lupus erythematosus and Grave's disease), inflammation, cancer  
CC (e.g. cancers of the bladder, brain, breast, colon and kidney,  
CC leukaemia), diseases of the nervous system and an infection of pathogenic  
CC organisms.  
XX  
SQ Sequence 51 BP; 20 A; 11 C; 11 G; 9 T; 0 other;  
Query Match 0.5%; Score 51; DB 22; Length 51;  
Best Local Similarity 100.0%; Pred. No. 0.0068;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9424 AAACCCATACCTAGAGACAAAGTCACTGAGACAAAGGCTTGAGGTCT 9474  
|||||  
Db 1 AAACCCATACCTAGAGACAAAGTCACTGAGACAAAGGCTTGAGGTCT 51  
RESULT 5  
AAL33289  
ID AAL33289 standard; DNA; 51 BP.  
XX  
XX AAL33289;  
AC  
XX  
XX 24-JAN-2002 (first entry)  
DE Human SNP oligonucleotide #5497.  
XX  
XX Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;  
KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;  
KW amyloid protein; angiotensin; apoptosis related protein; cadherin;  
KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;  
KW complement related protein; cytochrome; kinesin; cytokine; interferon;  
KW interleukin; G-protein coupled receptor; thioesterase; inflammation;  
KW multifactorial disease; autoimmune disease; infection;  
KW nervous system disease; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200147944-A2.  
XX  
XX 05-JUL-2001.  
XX  
XX 28-DEC-2000; 2000WO-US35498.  
XX  
XX 28-DEC-1999; 99US-0173419.  
XX 27-DEC-2000; 2000US-0173419.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Shimkets RA, Leach M;  
XX  
XX WPI; 2001-465210/50.  
XX  
XX Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,  
PT oncogenes and histones, useful for diagnosing and treating, e.g.  
PT cancer, autoimmune diseases and infections -  
XX Claim 1; Page 3241; 4143pp; English.  
XX  
XX The present invention relates to oligonucleotides encoding polymorphic  
CC variants of proteins related to amylases, amyloid proteins, angiotensin,  
CC

CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,  
 CC histones, kinases, colony stimulating factors, complement related  
 CC proteins, cytochromes, kinesins, cytokines, interferons, interleukins,  
 CC G-protein coupled receptors and thioesterases. The present sequence is  
 CC one such oligonucleotide. The oligonucleotides and the peptides encoded  
 CC by them may be used in the prevention, diagnosis and treatment of  
 CC diseases associated with inappropriate expression of the proteins listed  
 CC above. Disorders that may be prevented, diagnosed and/or treated include  
 CC multifactorial diseases with a genetic component, such as autoimmune  
 CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,  
 CC systemic lupus erythematosus and Grave's disease), inflammation, cancer  
 CC (e.g. cancers of the bladder, brain, breast, colon and kidney,  
 CC leukaemia), diseases of the nervous system and an infection of pathogenic  
 CC organisms.

SQ Sequence 51 BP; 22 A; 7 C; 13 G; 9 T; 0 other;  
 Query Match 0.5%; Score 51; DB 22; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 0.0068;  
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9712 GACAATGTGTGTCAAGAAATAACACCAAGTCTATAGGACAGTCAA 9762

Db 1 GACAATGTGTGTCAAGAAATAACACCAAGTCTATAGGACAGTCAA 51

## RESULT 6

AAL33290

ID AAL33290 standard; DNA; 51 BP.

XX AC

XX AAL33290;

XX DT

XX 24-JAN-2002 (first entry)

XX DE

XX Human SNP oligonucleotide #6498.

XX KW

KW Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;

KW neuroprotective; antimicrobial; gene therapy; vaccine; amylose; cancer;

KW amyloid protein; angiotensin; angiotensin II; angiotensin III; angiotensin IV;

KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;

KW complement related protein; cytochrome; kinesin; cytokine; interferon;

KW interleukin; G-protein coupled receptor; thioesterase; inflammation;

KW multifactorial disease; autoimmune disease; infection;

KW nervous system disease; ss.

XX OS

XX Homo sapiens.

XX PN

XX WO200147944-A2.

XX PD

XX 05-JUL-2001.

XX XX

XX 28-DEC-2000; 2000WO-US35498.

XX PF

XX 28-DEC-1999; 99US-0173419.

XX PR

XX 27-DEC-2000; 2000US-0173419.

XX XX

XX (CURA-) CURAGEN CORP.

XX XX

XX Shimkets RA, Leach M;

XX XX

XX WPI; 2001-465210/50.

XX XX

XX Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,  
 XX oncogenes and histones, useful for diagnosing and treating, e.g.  
 XX cancer, autoimmune diseases and infections -

XX PS

XX Claim 1; Page 3241; 4143pp; English.

XX XX

XX The present invention relates to oligonucleotides encoding polymorphic  
 XX variants of proteins related to amylases, amyloid proteins, angiotensin,  
 XX apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,  
 XX histones, kinases, colony stimulating factors, complement related  
 XX proteins, cytochromes, kinesins, cytokines, interferons, interleukins,

CC G-protein coupled receptors and thioesterases. The present sequence is  
 CC one such oligonucleotide. The oligonucleotides and the peptides encoded  
 CC by them may be used in the prevention, diagnosis and treatment of  
 CC diseases associated with inappropriate expression of the proteins listed  
 CC above. Disorders that may be prevented, diagnosed and/or treated include  
 CC multifactorial diseases with a genetic component, such as autoimmune  
 CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,  
 CC systemic lupus erythematosus and Grave's disease), inflammation, cancer  
 CC (e.g. cancers of the bladder, brain, breast, colon and kidney,  
 CC leukaemia), diseases of the nervous system and an infection of pathogenic  
 CC organisms.

SQ Sequence 51 BP; 20 A; 18 C; 7 G; 6 T; 0 other;  
 Query Match 0.5%; Score 51; DB 22; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 0.0068;  
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4840 GCCTCAATCTTCACACACACCTAGACAAAAACCCAGAGCTCCAAAG 4890

Db 1 GCCTCAATCTTCACACACACCTAGACAAAAACCCAGAGCTCCAAAG 51

## RESULT 7

AAH79977

ID AAH79977 standard; DNA; 51 BP.

XX XX

XX AC

XX AAH79977;

XX XX

XX 19-SEP-2001 (first entry)

XX DT

XX XX

XX DE

XX Human DNA containing single nucleotide polymorphism SEQ ID NO. 592.

XX XX

XX Human; single nucleotide polymorphism; SNP; angiotensin;  
 XX 4-hydroxybutyrate; dehydrogenase; protein therapy;XX adenosine triphosphate-dependent RNA helicase;  
 XX major histocompatibility complex Class I histocompatibility antigen; MHC;  
 XX phosphoglycerate kinase; immunosuppressive; immunostimulatory;  
 XX antirheumatic; antisclerotic; antidiabetic; antiinflammatory; cytostatic;  
 XX antileukemic; neuroprotective; antimicrobial; gene therapy; vaccine; ds.

XX KW

XX Homo sapiens.

XX OS

XX WO200148245-A2.

XX PN

XX 05-JUL-2001.

XX PD

XX 27-DEC-2000; 2000WO-US35346.

XX PF

XX 27-DEC-1999; 99US-0472688.

XX PR

XX (CURA-) CURAGEN CORP.

XX PA

XX Shimkets RA, Leach M;

XX PI

XX WPI; 2001-418297/44.

XX XX

XX Polymorphic nucleic acids encoding e.g. angiotensin, dehydrogenase,  
 XX adenosine triphosphate-dependent RNA helicase and/or phosphoglycerate  
 XX kinase, useful for diagnosing and treating, e.g. cancer, autoimmune  
 XX diseases and infections -

XX PS

XX Claim 1; Page 232; 484pp; English.

XX XX

XX The invention relates to nucleic acids (AAH79386-AAH80036) encoding  
 XX polymorphic variants of proteins (AAG98010-AAG98238) related to  
 XX angiotensin, 4-hydroxybutyrate, dehydrogenase, adenosine triphosphate  
 XX (ATP)-dependent RNA helicase, major histocompatibility complex (MHC)  
 XX Class I histocompatibility antigen and/or phosphoglycerate kinase. These  
 XX nucleic acid single nucleotide polymorphisms (SNPs) and the encoded  
 XX proteins have potential immunosuppressive, immunostimulatory,  
 XX antirheumatic, antisclerotic, antidiabetic, antiinflammatory, cytostatic,  
 XX antileukemic, neuroprotective and antimicrobial activity and may be

CC useful in gene/protein therapy, vaccines, modulation of the expression  
 CC and activity of proteins related to angiotensin, 4-hydroxybutyrate,  
 CC dehydrogenase, adenosine triphosphate (ATP)-dependent RNA helicase,  
 CC major histocompatibility complex (MHC) Class I histocompatibility antigen  
 CC and/or phosphoglycerate kinase. Disorders that may be prevented,  
 CC diagnosed and/or treated by the above methods include multifactorial  
 CC diseases with a genetic component, such as autoimmune diseases (e.g.  
 CC rheumatoid arthritis, multiple sclerosis, diabetes, systemic lupus  
 CC erythematosus and Grave's disease), inflammation, cancer (e.g. cancers  
 CC of the bladder, brain, breast, colon and kidney, leukemia), diseases of  
 CC the nervous system, an infection of pathogenic organisms. They may also  
 CC be used to alter phenotypic traits such as longevity, appearance,  
 CC strength, speed and endurance.

XX Sequence 51 BP; 15 A; 16 C; 12 G; 8 T; 0 other;

Query Match 0.5%; Score 51; DB 22; Length 51;

Best Local Similarity 100.0%; Pred. No. 0.0068;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4384 CAACCCCTAGAAGACCTGGCTGGTGAAGAGCTCTCCAGACACCACTA 4434

Db 1 CAACCCCTAGAAGACCTGGCTGGTGAAGAGCTCTCCAGACACCACTA 51

# RESULT 8

AAL27415

ID AAL27415 standard; DNA; 51 BP.

XX

AC AAL27415;

XX

DT 24-JAN-2002 (first entry)

XX

DE Human SNP oligonucleotide #623.

XX

KW Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;  
 KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;  
 KW amyloid protein; angiotensin; apoptosis related protein; cadherin;  
 KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;  
 KW complement related protein; cytochrome; kinesin; cytokine; interferon;  
 KW interleukin; G-protein coupled receptor; thioesterase; inflammation;  
 KW multifactorial disease; autoimmune disease; infection;  
 KW nervous system disease; ss.

XX Homo sapiens.

XX

PN WO200147944-A2.

XX

PD 05-JUL-2001.

XX

PF 28-DEC-2000; 2000WO-US35498.

XX

PR 28-DEC-1999; 99US-0173419.

XX

PR 27-DEC-2000; 2000US-0173419.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Shinkets RA, Leach M;

XX

DR WPI; 2001-465210/50.

XX

PT Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,  
 PT oncogenes and histones, useful for diagnosing and treating, e.g.  
 PT cancer, autoimmune diseases and infections.

XX

PS Claim 1; Page 1555; 4143pp; English.

XX

CC The present invention relates to oligonucleotides encoding polymorphic  
 CC variants of proteins related to amylases, amyloid proteins, angiotensin,  
 CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,  
 CC histones, kinases, colony stimulating factors, complement related  
 CC proteins, cytochromes, kinesins, cytokines, interferons, interleukins,  
 CC G-protein coupled receptors and thioesterases. The present sequence is

CC one such oligonucleotide. The oligonucleotides and the peptides encoded  
 CC by them may be used in the prevention, diagnosis and treatment of  
 CC diseases associated with inappropriate expression of the proteins listed  
 CC above. Disorders that may be prevented, diagnosed and/or treated include  
 CC multifactorial diseases with a genetic component, such as autoimmune  
 CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,  
 CC systemic lupus erythematosus and Grave's disease), inflammation, cancer  
 CC (e.g. cancers of the bladder, brain, breast, colon and kidney,  
 CC leukemia), diseases of the nervous system and an infection of pathogenic  
 CC organisms.

XX Sequence 51 BP; 23 A; 11 C; 9 G; 8 T; 0 other;

Query Match 0.5%; Score 49.4; DB 22; Length 51;

Best Local Similarity 98.0%; Pred. No. 0.017;

Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9221 AGAGGACACAAATTTACAGACTCGTCCCTGAAATAGGGAATATCCC 9271

Db 1 AAGAGGACACAAATTTACAGACTCGTCCCTGAAATAGGGAATATCCC 51

# RESULT 9

AAL33291

ID AAL33291 standard; DNA; 51 BP.

XX

AC AAL33291;

XX

DT 24-JAN-2002 (first entry)

XX

DE Human SNP oligonucleotide #6499.

XX

KW Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;  
 KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;  
 KW amyloid protein; angiotensin; apoptosis related protein; cadherin;  
 KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;  
 KW complement related protein; cytochrome; kinesin; cytokine; interferon;  
 KW interleukin; G-protein coupled receptor; thioesterase; inflammation;  
 KW multifactorial disease; autoimmune disease; infection;  
 KW nervous system disease; ss.

XX Homo sapiens.

XX

PN WO200147944-A2.

XX

PD 05-JUL-2001.

XX

PF 28-DEC-2000; 2000WO-US35498.

XX

PR 28-DEC-1999; 99US-0173419.

XX

PR 27-DEC-2000; 2000US-0173419.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Shinkets RA, Leach M;

XX

DR WPI; 2001-465210/50.

XX

PT Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,  
 PT oncogenes and histones, useful for diagnosing and treating, e.g.  
 PT cancer, autoimmune diseases and infections.

XX

PS Claim 1; Page 3241; 4143pp; English.

XX

CC The present invention relates to oligonucleotides encoding polymorphic  
 CC variants of proteins related to amylases, amyloid proteins, angiotensin,  
 CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,  
 CC histones, kinases, colony stimulating factors, complement related  
 CC proteins, cytochromes, kinesins, cytokines, interferons, interleukins,  
 CC G-protein coupled receptors and thioesterases. The present sequence is  
 CC one such oligonucleotide. The oligonucleotides and the peptides encoded  
 CC by them may be used in the prevention, diagnosis and treatment of  
 CC diseases associated with inappropriate expression of the proteins listed

CC above. Disorders that may be prevented, diagnosed and/or treated include  
 CC multifactorial diseases with a genetic component, such as autoimmune  
 CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,  
 CC systemic lupus erythematosus and Grave's disease), inflammation, cancer  
 CC (e.g. cancers of the bladder, brain, breast, colon and kidney,  
 CC leukaemia), diseases of the nervous system and an infection of pathogenic  
 CC organisms.

SQ Sequence 51 BP; 22 A; 11 C; 13 G; 5 T; 0 other;

Query Match 0.5%; Score 49.4; DB 22; Length 51;  
 Best Local Similarity 98.0%; Pred. No. 0.017;  
 Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7798 ACAAGAGATGCCCAAGACAGTCCAGGAAGAGTAAAGAGAGGCTC 7848

Db 1 ACAAGAGATGCCCAAGACAGTCCAGGAAGAGTAAAGAGAGGCTC 51

RESULT 10

AAL34120

ID AAL34120 standard; DNA; 51 BP.

XX AC AAL34120;

XX AC AAL34120;

XX 24-JAN-2002 (first entry)

XX Human SNP oligonucleotide #7328.

XX Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;  
 KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;  
 KW amyloid protein; angioleiectin; apoptosis related protein; cadherin;  
 KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;  
 KW complement related protein; cytochrome; kinesin; cytokine; interferon;  
 KW interleukin; G-protein coupled receptor; thioesterase; inflammation;  
 KW multifactorial disease; autoimmune disease; infection;  
 KW nervous system disease; ss.

XX Homo sapiens.

XX WO200147944-A2.

XX 05-JUL-2001.

XX 28-DEC-2000; 2000WO-US35498.

XX 28-DEC-1999; 99US-0173419.

XX 27-DEC-2000; 2000US-0173419.

XX (CURA-) CURAGEN CORP.

XX ShImkets RA, Leach M;

XX WPI; 2001-465210/50.

XX Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,  
 PT oncogenes and histones, useful for diagnosing and treating, e.g.  
 PT cancer, autoimmune diseases and infections -

XX Claim 1; Page 3494; 4143pp; English.

XX The present invention relates to oligonucleotides encoding polymorphic  
 CC variants of proteins related to amylases, amyloid proteins, angioleiectin,  
 CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,  
 CC histones, kinases, colony stimulating factors, complement related  
 CC proteins, cytochromes, kinesins, cytokines, interferons, interleukins,  
 CC G-protein coupled receptors and thioesterases. The present sequence is  
 CC one such oligonucleotide. The oligonucleotides and the peptides encoded  
 CC by them may be used in the prevention, diagnosis and treatment of  
 CC diseases associated with inappropriate expression of the proteins listed  
 CC above. Disorders that may be prevented, diagnosed and/or treated include  
 CC multifactorial diseases with a genetic component, such as autoimmune  
 CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,

CC systemic lupus erythematosus and Grave's disease), inflammation, cancer  
 CC (e.g. cancers of the bladder, brain, breast, colon and kidney,  
 CC leukaemia), diseases of the nervous system and an infection of pathogenic  
 CC organisms.

SQ Sequence 51 BP; 20 A; 9 C; 6 G; 16 T; 0 other;

Query Match 0.5%; Score 49.4; DB 22; Length 51;  
 Best Local Similarity 98.0%; Pred. No. 0.017;  
 Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 673 CAATGCTTGACAATAGCAAAAAAATGAATCTCCCTTTTGAAGCTTTAT 723

Db 1 CAATGCTTGACAATAGCAAAAAAATGAATCTCCCTTTTGAAGCTTTAT 51

RESULT 11

ABN51074

ID ABN51074 standard; DNA; 65 BP.

XX AC ABN51074;

XX AC ABN51074;

XX 15-JUL-2002 (first entry)

XX Mouse spliced transcript detection oligonucleotide SEQ ID NO:23822.

XX Human; mouse; rat; splice transcript; detection; RNA transcript;  
 KW splice variant; transcriptome; oligonucleotide library; ss.

XX Mus musculus.

XX WO200210449-A2.

XX 07-FEB-2002.

XX 20-JUL-2001; 2001WO-IB01903.

XX 28-JUL-2000; 2000US-221607P.

XX 02-MAY-2001; 2001US-287724P.

XX (COMP-) COMPUGEN INC.

XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;

XX WPI; 2002-257383/30.

XX New oligonucleotide libraries comprising oligonucleotides which  
 PT selectively hybridize to mRNAs transcribed from a transcription unit of  
 PT a genome, useful for detecting tissue-, pathology-, and  
 PT developmental-specific genes -

XX Example 1; SEQ ID 23822; 47pp; English.

XX The present invention describes oligonucleotide libraries for detecting  
 CC messenger RNAs that populate a (sub-)transcriptome, where the  
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
 CC transcription units that populate a genome. The library comprises  
 CC several oligonucleotides, each capable of hybridizing selectively to a  
 CC set of messenger RNAs transcribed from a given transcription unit of  
 CC the genome, which encodes one or more messenger RNA splice variants.  
 CC The oligonucleotide libraries are useful for detecting mRNAs from a  
 CC biological sample, in expression profiling studies, in qualitatively or  
 CC quantitatively characterising the corresponding transcriptome, and in  
 CC detecting RNA transcripts and splice variants of human or animal  
 CC transcriptomes. The libraries may also be used as specialised mini  
 CC libraries to detect transcripts of a sub-transcriptome under a  
 CC particular biological or pathological state, and so allowing the  
 CC detection of tissue- and pathology-specific genes such as those genes  
 CC only expressed in specific tissue under a specific pathological  
 CC condition; to detect developmental specific genes; and to detect RNA  
 CC transcripts and splice variants of a transcriptome of a patient suffering  
 CC from a particular disorder. ABN27253 to ABN59589 represent  
 CC oligonucleotide sequences from rats, humans and mice, which are used in

Ö

```

XX Melanoma; thyroid tumour; MEL3; MEL7; Thy5; Thy6; Thy11; Thy14; Thy15;
KW rectal cancer; lung cancer; breast cancer; colon cancer; PCR primer;
KW Ki-67 nuclear antigen gene; ss.
XX Homo sapiens.
XX
XX WO200050595-A2.
XX
XX
XX PD
XX 31-AUG-2000.
XX
XX PF 25-FEB-2000; 2000WO-US04929.
XX
XX PR 25-FEB-1999; 99US-0257417.
XX
XX PA (GOUT/) GOUT I.
XX PA (RODN/) RODNIN N.
XX PA (FILO/) FILONENKO V.
XX PA (MATS/) MATSUKA G.
XX PA (SCAN/) SCANLAN M.
XX PA (OLDL/) OLD L.
XX PA (BILY/) BILYNSKY B.
XX
XX PI Gout I, Rodnin N, Filonenko V, Matsuka G, Scanlan M, Old L;
XX PI Bilynsky B;
XX
XX DR WPI; 2000-572092/53.
XX
XX PT Novel isolated nucleic acid molecules for diagnosing and treating
XX PT melanoma, thyroid tumors, rectal, lung, breast and colon cancers -
XX
XX PS Example 3; Page 48; 94pp; English.
XX
XX CC The specification describes polynucleotides which are associated with
XX CC melanoma and thyroid tumors. Specifically, these are MEL3, MEL7, Thy5,
XX CC Thy6, Thy11, Thy14, and Thy15. The polynucleotides are useful for
XX CC diagnosing and treating a patient with melanoma, thyroid tumour,
XX CC rectal cancer, lung cancer, breast cancer or colon cancer. PCR
XX CC primers AAA64666-67 were used to amplify a fragment of MEL-9 clone,
XX CC which encodes a Ki-67 nuclear antigen gene, which is a polynucleotide
XX CC of the invention.
XX
XX SQ Sequence 53 BP; 18 A; 21 C; 8 G; 6 T; 0 other;
XX
XX Query Match 0.3%; Score 27.8; DB 21; Length 53;
XX Best Local Similarity 93.5%; Pred. No. 4.4e+03;
XX Matches 29; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4316 AACAGAACTGGACCCAGCAGCAAGTGTAAAC 4346
DB 23 ACCACAACCTGGACCCAGCAGCAAGTGTAAAC 53

RESULT 15
AAA64667/c
ID AAA64667 standard; DNA; 36 BP.
XX
XX AC AAA64667;
XX
XX DT 02-JAN-2001 (first entry)
XX
XX DE PCR primer used to amplify a 500 bp fragment of a MEL9 clone.
XX
XX KW Melanoma; thyroid tumour; MEL3; MEL7; Thy5; Thy6; Thy11; Thy14; Thy15;
XX KW rectal cancer; lung cancer; breast cancer; colon cancer; PCR primer;
XX KW Ki-67 nuclear antigen gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200050595-A2.
XX
XX PD 31-AUG-2000.
XX

```

```

PF 25-FEB-2000; 2000WO-US04929.
XX
XX PR 25-FEB-1999; 99US-0257417.
XX
XX PA (GOUT/) GOUT I.
XX PA (RODN/) RODNIN N.
XX PA (FILO/) FILONENKO V.
XX PA (MATS/) MATSUKA G.
XX PA (SCAN/) SCANLAN M.
XX PA (OLDL/) OLD L.
XX PA (BILY/) BILYNSKY B.
XX
XX PI Gout I, Rodnin N, Filonenko V, Matsuka G, Scanlan M, Old L;
XX PI Bilynsky B;
XX
XX DR WPI; 2000-572092/53.
XX
XX PT Novel isolated nucleic acid molecules for diagnosing and treating
XX PT melanoma, thyroid tumors, rectal, lung, breast and colon cancers -
XX
XX PS Example 3; Page 48; 94pp; English.
XX
XX CC The specification describes polynucleotides which are associated with
XX CC melanoma and thyroid tumors. Specifically, these are MEL3, MEL7, Thy5,
XX CC Thy6, Thy11, Thy14, and Thy15. The polynucleotides are useful for
XX CC diagnosing and treating a patient with melanoma, thyroid tumour,
XX CC rectal cancer, lung cancer, breast cancer or colon cancer. PCR
XX CC primers AAA64666-67 were used to amplify a fragment of MEL-9 clone,
XX CC which encodes a Ki-67 nuclear antigen gene, which is a polynucleotide
XX CC of the invention.
XX
XX SQ Sequence 36 BP; 7 A; 10 C; 8 G; 11 T; 0 other;
XX
XX Query Match 0.3%; Score 27.4; DB 21; Length 36;
XX Best Local Similarity 96.6%; Pred. No. 4.6e+03;
XX Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7279 CCTAAGGAAAGGCTGAGGCTCTAGAGGA 7307
DB 36 CCTAAGGAAAGGCTGAGGCTCTATAGGA 8

Search completed: December 3, 2002, 02:19:34
Job time : 1189 secs

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2002, 22:44:20 ; Search time 191 Seconds  
(without alignments)  
15680.656 Million cell used

Title: US-09-700-906A-1\_COPY\_197\_9962  
 Perfect score: 9766  
 Sequence: 1 atgtatgccacagagagcct.....tcataggacactgaagata 9766

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 644616

```
Minimum DB seq length: 0
Maximum DB seq length: 66
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 s
```

```
Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	25	0.3	50	1	US-08-171-389-400	Sequence 400, App
2	25	0.3	50	1	US-08-123-936-400	Sequence 400, App
3	25	0.3	50	2	US-08-475-228A-400	Sequence 400, App
4	25	0.3	50	3	US-08-482-080A-400	Sequence 400, App
5	25	0.3	50	4	US-09-354-947-400	Sequence 400, App
6	25	0.3	50	5	PCR-US93-12388-400	Sequence 400, App
C 7	24.8	0.3	50	1	US-08-207-901-43	Sequence 43, Appl
	24.8	0.3	66	4	US-09-097-319A-48	Sequence 48, Appl
8	24.2	0.2	62	1	US-08-222-177A-56	Sequence 56, Appl
9	24.2	0.2	62	1	US-08-153-071-13	Sequence 13, Appl
10	24	0.2	54	1	US-08-609-271-20	Sequence 20, Appl
11	24	0.2	54	1	US-08-438-511-13	Sequence 13, Appl
12	24	0.2	54	2	US-08-487-431-16	Sequence 16, Appl
13	24	0.2	54	2	US-08-188-374-20	Sequence 20, Appl
14	24	0.2	54	3	US-08-973-629-16	Sequence 16, Appl
15	24	0.2	54	3	US-08-910-632-29	Sequence 29, Appl
C 16	24	0.2	58	3	US-08-805-631A-29	Sequence 29, Appl
C 17	24	0.2	58	3	US-09-569-344-29	Sequence 29, Appl
C 18	24	0.2	58	4	US-08-222-177A-169	Sequence 169, App
19	23.4	0.2	50	1	US-08-222-177A-282	Sequence 282, App
20	23.4	0.2	61	1	US-08-477-527A-67	Sequence 67, Appl
C 21	23.2	0.2	46	2	US-08-481-710-67	Sequence 67, Appl
C 22	23.2	0.2	46	3	PCR-US96-09537-67	Sequence 67, Appl
C 23	23.2	0.2	46	5	US-08-273-594-33	Sequence 33, Appl
C 24	23.2	0.2	57	1	US-09-094-919-33	Sequence 33, Appl
C 25	23.2	0.2	57	3	US-08-222-177A-256	Sequence 256, App
26	23	0.2	51	1	US-08-455-627-22	Sequence 22, Appl
27	22.8	0.2	42	1		

## ALIGNMENTS

RESULT 1  
 US-08-171-389-400  
 ; Sequence 400, Application US/08171389  
 ; Patent No. 5578444  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edwards, Cynthia A.  
 ; APPLICANT: Cantor, Charles R.  
 ; APPLICANT: Andrews, Beth M.  
 ; APPLICANT: Turin, Lisa M.  
 ; APPLICANT: Fry, Kirk E.  
 ; TITLE OF INVENTION: Sequence-Directed DNA Binding  
 ; TITLE OF INVENTION: Molecules, Compositions and Methods  
 ; NUMBER OF SEQUENCES: 641  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genelabs Technologies, Inc.  
 ; STREET: 505 Penobscot Drive  
 ; CITY: Redwood City  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94063  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/171,389  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/123,936  
 ; FILING DATE: 17-SEP-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/996,783  
 ; FILING DATE: 23-DEC-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/723,618  
 ; FILING DATE: 27-JUN-1991  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/081,070  
 ; FILING DATE: 22-JUN-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fabian, Gary R.  
 ; REGISTRATION NUMBER: 33,875  
 ; REFERENCE/DOCKET NUMBER: 4600-0175/G19P3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 324-0880  
 ; TELEFAX: (415) 324-0960  
 ; INFORMATION FOR SEQ ID NO: 400:  
 ; SEQUENCE CHARACTERISTICS:



QY 6062 CGTCAGGGAAGACCACACAGACACACAGAGAGACAGACAGCA 6102  
|| ||| ||||| | | ||||| ||||| ||||| |  
Db 9 CGGCAGTGAAGATAAAGGCCACACAGAGAGCGCAGCA 49

## RESULT 4

US-08-482-080A-400  
; Sequence 400, Application US/08482080A  
; Patent No. 6010849  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Cynthia A.  
; APPLICANT: Cantor, Charles R.  
; APPLICANT: Andrews, Beth M.  
; APPLICANT: Turin, Lisa M.  
; APPLICANT: Fry, Kirk E.  
; TITLE OF INVENTION: Sequence-Directed DNA Binding  
; TITLE OF INVENTION: Molecules, Compositions and Methods  
; NUMBER OF SEQUENCES: 664  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,080A  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/171,389  
; FILING DATE: 20-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/123,936  
; FILING DATE: 17-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/996,783  
; FILING DATE: 23-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/723,618  
; FILING DATE: 27-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/081,070  
; FILING DATE: 22-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brady, John F.  
; REGISTRATION NUMBER: 39,118  
; REFERENCE/DOCKET NUMBER: 4600-0175.20/G19P3D1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 324-0880  
; TELEFAX: (650) 324-0960  
; INFORMATION FOR SEQ ID NO: 400:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Orangutan epsilon-globin gene with  
; INDIVIDUAL ISOLATE: flanking Alu repeats  
US-08-482-080A-400

Query Match 0.3%; Score 25; DB 3; Length 50;  
Best Local Similarity 75.6%; Pred. No. 1.1e+03;  
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 6062 CGTCAGGGAAGACCACACAGACACACAGAGAGACAGACAGCA 6102  
|| ||| ||||| | | ||||| ||||| ||||| |  
Db 9 CGGCAGTGAAGATAAAGGCCACACAGAGAGCGCAGCA 49

## RESULT 5

US-09-354-947-400  
; Sequence 400, Application US/09354947  
; Patent No. 6384208  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Cynthia A.  
; APPLICANT: Cantor, Charles R.  
; APPLICANT: Andrews, Beth M.  
; APPLICANT: Turin, Lisa M.  
; APPLICANT: Fry, Kirk E.  
; TITLE OF INVENTION: Sequence-Directed DNA Binding  
; TITLE OF INVENTION: Molecules, Compositions and Methods  
; NUMBER OF SEQUENCES: 664  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/354,947  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/482,080  
; FILING DATE: 07-JUN-1995  
; APPLICATION NUMBER: US 08/171,389  
; FILING DATE: 20-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/123,936  
; FILING DATE: 17-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/996,783  
; FILING DATE: 23-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/723,618  
; FILING DATE: 27-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/081,070  
; FILING DATE: 22-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brady, John F.  
; REGISTRATION NUMBER: 39,118  
; REFERENCE/DOCKET NUMBER: 4600-0175.20/G19P3D1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 324-0880  
; TELEFAX: (650) 324-0960  
; INFORMATION FOR SEQ ID NO: 400:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Orangutan epsilon-globin gene with  
; INDIVIDUAL ISOLATE: flanking Alu repeats  
US-09-354-947-400

Query Match 0.3%; Score 25; DB 4; Length 50;  
Best Local Similarity 75.6%; Pred. No. 1.1e+03;  
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
; QY 6062 CGTCAGGGAAGACCACACAGACACACAGACAGACAGCA 6102
; ||| ||||| | | ||||| ||||| ||||| |
; Db 9 CGGCAGTGAAGATAAAGGCCACACAGAGAGGCAGCAGCA 49

RESULT 6
PCT-US93-12388-400
; Sequence 400, Application PC/TUS9312388
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 641
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12388
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0175.41/G19PCT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 400:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Orangutan epsilon-globin gene with
; INDIVIDUAL ISOLATE: flanking Alu repeats
PCT-US93-12388-400

Query Match 0.3%; Score 25; DB 5; Length 50;
Best Local Similarity 75.6%; Pred. No. 1.1e+03;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

; QY 6062 CGTCAGGGAAGACCACACAGACACACAGACAGACAGCA 6102
; ||| ||||| | | ||||| ||||| ||||| |
; Db 9 CGGCAGTGAAGATAAAGGCCACACAGAGAGGCAGCAGCA 49

RESULT 7
US-08-207-901-43/c
; Sequence 43, Application US/08207901
; Patent No. 5629153
; GENERAL INFORMATION:
; APPLICANT: Urdea, Michael S.
; TITLE OF INVENTION: USE OF DNA-DEPENDENT RNA POLYMERASE
; TITLE OF INVENTION: TRANSCRIPTS AS REPORTER MOLECULES FOR SIGNAL
```

```
; TITLE OF INVENTION: AMPLIFICATION IN NUCLEIC ACID HYBRIDIZATION ASSAYS
; NUMBER OF SEQUENCES: 120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,901
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/639,560B
; FILING DATE: 10-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Kenneth M.
; REGISTRATION NUMBER: 34,174
; REFERENCE/DOCKET NUMBER: 00081.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2719
; TELEFAX: 510-655-3542
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-207-901-43

Query Match 0.3%; Score 24.8; DB 1; Length 50;
Best Local Similarity 69.6%; Pred. No. 1.2e+03;
Matches 32; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

; QY 2073 ACCCTGCTACTCCAAAGAGCCTGTGGCGAAGTTTCACAGTCAATTT 2118
; ||| | ||||| ||||| ||||| | ||||| |||
; Db 49 ACCACTTTCCTCCAAAGAGCCTATGGGAGKGGGCGCTCAGYCGGTTT 4

RESULT 8
US-09-097-319A-48
; Sequence 48, Application US/09097319A
; Patent No. 6384,207
; GENERAL INFORMATION:
; APPLICANT: Ainley, Michael
; APPLICANT: Armstrong, Katherine
; APPLICANT: Belmar, Scott
; APPLICANT: Folkerts, Otto
; APPLICANT: Hopkins, Nicole
; APPLICANT: Menke, Michael A.
; APPLICANT: Paredy, Dayakar
; APPLICANT: Petolino, Joseph F.
; APPLICANT: Smith, Kelley
; APPLICANT: Woosley, Aaron
; TITLE OF INVENTION: Regulatory Sequences for Transgenic Plants
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dowelanco Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/097.319A  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stuart, Donald R  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317 337 4816  
; TELEFAX: 317 337 4847  
; INFORMATION FOR SEQ ID NO: 48:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 66 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
US-09-097-319A-48

Query Match 0.3%; Score 24.8; DB 4; Length 66;  
Best Local Similarity 63.3%; Pred. No. 1.5e+03;  
Matches 38; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 7789 GCCACGACACAAAGAGATGCCCCAAAGACACGCTCCAGGAAGAAGTAAAGAGAGAGCTC 7848  
||||| || ||| || ||||| || ||||| ||| || ||| || |||||  
Db 5 GCGGATAGTGGAAACCGAGCGCCACGACACTGTCGAGGCGCAAGAAATAGTAAGAGCTC 64

RESULT 9  
US-08-222-177A-56  
; Sequence 56, Application US/08222177A  
; Patent No. 5582979  
; GENERAL INFORMATION:  
; APPLICANT: Weber, James L.  
; TITLE OF INVENTION: LENGTH POLYMORPHISMS IN  
; TITLE OF INVENTION: (dC-dA)n.(dG-dT)n SEQUENCES AND METHODS OF USING SAME  
; NUMBER OF SEQUENCES: 460  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dewitt Ross & Stevens, S.C.  
; STREET: 8000 Excelsior Drive, Suite 401  
; CITY: Madison  
; STATE: Wisconsin  
; COUNTRY: USA  
; ZIP: 53717-1914  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/222.177A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/341.562  
; FILING DATE: 21-APR-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sara, Charles S.  
; REGISTRATION NUMBER: 30,492  
; REFERENCE/DOCKET NUMBER: 09865.601  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (608) 831-2100  
; TELEFAX: (608) 831-2106  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 56:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 62 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)

; IMMEDIATE SOURCE:  
; CLONE: mfd2rs  
US-08-222-177A-56  
Query Match 0.2%; Score 24.2; DB 1; Length 62;  
Best Local Similarity 71.1%; Pred. No. 2.1e+03;  
Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 4954 CTCACACACATCAGGAGACTACACACACACACACACACACACACACACCA 4998  
||||| ||| || || ||||| ||||| ||| ||  
Db 4 CA 48  
RESULT 10  
US-08-153-071-13  
; Sequence 13, Application US/08153071  
; Patent No. 5665869  
; GENERAL INFORMATION:  
; APPLICANT: Ryland, James R.  
; APPLICANT: Matthews, Maura-Ann H.  
; APPLICANT: Ernst, Ulrich P.  
; APPLICANT: Houk, Daniel E.  
; APPLICANT: Travlor, David W.  
; APPLICANT: Williams, Lee R.  
; APPLICANT: Mitchell, David J.  
; APPLICANT: Chivers, Mark L.  
; APPLICANT: Belval, Thomas K.  
; TITLE OF INVENTION: Method for the Rapid Removal of  
; TITLE OF INVENTION: Protoporphylin IX from Protoporphylin IX-Containing Solutio  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Somatogen, Inc.  
; STREET: 5797 Central Avenue  
; CITY: Boulder  
; STATE: Colorado  
; ZIP: 80301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: System 7.0.1  
; SOFTWARE: Microsoft Word 5.0a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/153.071  
; FILING DATE: No. 5665869ember 15, 1993  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5665869ak, Henry P.  
; REGISTRATION NUMBER: 33200  
; REFERENCE/DOCKET NUMBER: 120  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303-541-3322  
; TELEFAX: 303-444-3013  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 54  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: primer for pBR322 ori positions  
; MOLECULE TYPE: 3170-3148  
; HYPOTHETICAL: no  
US-08-153-071-13

Query Match 0.2%; Score 24; DB 1; Length 54;  
Best Local Similarity 68.8%; Pred. No. 2.2e+03;  
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
QY 31 AAAAGAGCGGGTCGACGGTCCCGACCTTCCCGCTGAGCTCAGCACC 78  
||||| || || ||| || || ||||| ||||| ||  
Db 6 AAAGATCCAAGTAGCCGCGCGCGTTCACCTGAGGCTCAGACCC 53

RESULT 11

```
US-08-609-271-20
; Sequence 20, Application US/08609271
; Patent No. 5811264
; GENERAL INFORMATION:
; APPLICANT: Aitken, Jacqueline F.
; APPLICANT: Apostol, Izidor Z.
; APPLICANT: Lippincott, Julie A.
; APPLICANT: Levine, Joseph D.
; TITLE OF INVENTION: Proteins with Mutations to Decrease N-Terminal Methylation
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Somatogen, Inc.
; STREET: 2545 Central Avenue, Site FD-1
; CITY: Boulder
; STATE: Colorado
; ZIP: 80301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.0.1
; SOFTWARE: Microsoft Word 5.0a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,271
; FILING DATE: 28 February 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/188,374
; FILING DATE: 1/27/94
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5811264elli, Marianne F.
; REGISTRATION NUMBER: 38571
; NAME: Brown, Theresa A.
; REGISTRATION NUMBER: 32547
; REFERENCE/DOCKET NUMBER: 170/Div
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-541-3324
; TELEFAX: 303-444-3013
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: primer for pBR322 ori positions 3170-3148
; HYPOTHETICAL: no
US-08-609-271-20

Query Match 0.2%; Score 24; DB 1; Length 54;
Best Local Similarity 68.8%; Pred. No. 2.2e+03;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 31 AAAAGGAGCGGGTCGACGGTCCCGACCTTCCCGCTGAGCGCTCAGCACCC 78
Db 6 AAAGGATCCAAGTAGCGCGCGCGCGCTTCCACTGAGCGCTCAGCACCC 53

RESULT 12
US-08-438-511-13
; Sequence 13, Application US/08438511
; Patent No. 5840851
; GENERAL INFORMATION:
; APPLICANT: Plomer, J. Jeffrey
; APPLICANT: Ryland, James R.
; APPLICANT: Matthews, Maura-Ann H.
; APPLICANT: Traylor, David W.
; APPLICANT: Milne, Erin E.
; APPLICANT: Durfee, Steven L.
; APPLICANT: Mathews, Antony J.
; TITLE OF INVENTION: Purification of Hemoglobin
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Somatogen, Inc.
```

```
; STREET: 2545 Central Avenue, Suite FDI
; CITY: Boulder
; STATE: Colorado
; ZIP: 80301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.0.1
; SOFTWARE: Microsoft Word 5.0a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/438,511
; FILING DATE: 10 May 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/339,304
; FILING DATE: No. 5840851ember 14, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/097,273
; FILING DATE: July 23, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5840851elli, Marianne F.
; REGISTRATION NUMBER: 38571
; REFERENCE/DOCKET NUMBER: 191
; NAME: Brown, Theresa A.
; REGISTRATION NUMBER: 32,547
; REFERENCE/DOCKET NUMBER: 191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-541-3324
; TELEFAX: 303-444-3013
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: primer for pBR322 ori positions 3170-3148
; HYPOTHETICAL: no
US-08-438-511-13

Query Match 0.2%; Score 24; DB 2; Length 54;
Best Local Similarity 68.8%; Pred. No. 2.2e+03;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 31 AAAAGGAGCGGGTCGACGGTCCCGACCTTCCCGCTGAGCGCTCAGCACCC 78
Db 6 AAAGGATCCAAGTAGCGCGCGCGCGCTTCCACTGAGCGCTCAGCACCC 53

RESULT 13
US-08-487-431-16
; Sequence 16, Application US/08487431
; Patent No. 5844090
; GENERAL INFORMATION:
; APPLICANT: Anderson, David C.
; APPLICANT: Mathews, Antony J.
; APPLICANT: Trimble, Stephen P.
; APPLICANT: Anthony-Cahill, Spencer
; TITLE OF INVENTION: Modified Hemoglobin-like Compounds
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Somatogen, Inc.
; STREET: 2545 Central Avenue, Suite FDI
; CITY: Boulder
; STATE: Colorado
; ZIP: 80301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: Microsoft Word 5.0a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,431
```

; FILING DATE: June 7, 1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/240,712  
; FILING DATE: No. 5844090ember 6, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brown, Theresa A.  
; REGISTRATION NUMBER: 32,547  
; REFERENCE/DOCKET NUMBER: 61  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ramsey R. Stewart  
; REGISTRATION NUMBER: 38,322  
; REFERENCE/DOCKET NUMBER: 61  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303-541-3356  
; TELEFAX: 303-444-3013  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 54  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: primer for pBR322 ori positions 3170-3148  
; HYPOTHETICAL: no  
US-08-487-431-16

Query Match 0.2%; Score 24; DB 2; Length 54;  
Best Local Similarity 68.8%; Pred. No. 2.2e+03;  
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
QY 31 AAAGGAGGGGGTCGAGCGTCCACACTTCCCTGAGCCTCAGCACC 78  
||||| | | | | | | | | | | | | | | | | | | | |  
Db 6 AAAGGATCCAAAGTAGCGCGCGCGTTCACACTGAGCGTCAGACCC 53

RESULT 14  
US-08-188-374-20  
; Sequence 20, Application US/08188374B  
; Patent No. 6140071  
; GENERAL INFORMATION:  
; APPLICANT: Aitken F., Jacqueline  
; APPLICANT: Apostol, Izydor Z.  
; APPLICANT: Lippincott, Julie A.  
; APPLICANT: Levine, Joseph D.  
; TITLE OF INVENTION: Proteins with Mutations to Decrease N-Terminal Meth  
; FILE REFERENCE: BXTB 1953  
; CURRENT APPLICATION NUMBER: US/08/188,374B  
; CURRENT FILING DATE: 1994-01-27  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 54  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: primer for ori  
US-08-188-374-20

Query Match 0.2%; Score 24; DB 3; Length 54;  
Best Local Similarity 68.8%; Pred. No. 2.2e+03;  
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
QY 31 AAAGGAGGGGGTCGAGCGTCCACACTTCCCTGAGCCTCAGCACC 78  
||||| | | | | | | | | | | | | | | | | | | | |  
Db 6 AAAGGATCCAAAGTAGCGCGCGCGTTCACACTGAGCGTCAGACCC 53

RESULT 15  
US-08-973-629-16  
; Sequence 16, Application US/08973629A  
; Patent No. 6150506  
; GENERAL INFORMATION:

; APPLICANT: Trimble, Stephen  
; APPLICANT: Mathews, Anthony  
; APPLICANT: Kerwin, Bruce  
; APPLICANT: Marquardt, David  
; APPLICANT: Anthony-Cahill, Spencer  
; APPLICANT: Epp, Janice  
; APPLICANT: Madril, Dominic  
; APPLICANT: Anderson, David  
; TITLE OF INVENTION: MODIFIED HEMOGLOBIN-LIKE COMPOUNDS AND METHODS OF  
; ATTORNEY/AGENT INFORMATION:  
; FILE REFERENCE: BXTB 1928  
; CURRENT APPLICATION NUMBER: US/08/973,629A  
; CURRENT FILING DATE: 1998-08-24  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 16  
; LENGTH: 54  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: Oligonucleotide or 5' primer for production of a  
; OTHER INFORMATION: pBR322 ori with Bam HI and No. 6150506 I sites  
US-08-973-629-16

Query Match 0.2%; Score 24; DB 3; Length 54;  
Best Local Similarity 68.8%; Pred. No. 2.2e+03;  
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
QY 31 AAAGGAGGGGGTCGAGCGTCCACACTTCCCTGAGCCTCAGCACC 78  
||||| | | | | | | | | | | | | | | | | | | | |  
Db 6 AAAGGATCCAAAGTAGCGCGCGCGTTCACACTGAGCGTCAGACCC 53

Search completed: December 3, 2002, 02:23:04  
Job time : 199 secs

**THIS PAGE BLANK (USPTO)**





```
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946.807
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795.668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515.716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1349
; LENGTH: 61
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-946-807-1349

Query Match          0.2%; Score 23.4; DB 9; Length 61;
Best Local Similarity 61.0%; Pred. No. 1.2e+04;
Matches 36; Conservative 1; Mismatches 22; Indels 0; Gaps 0;

QY   2233 AAATGGACCTTTAGGGAAGATCTTTCAGGATAGCTGAAATGTTCAGAACCCCAAGTGAA 2291
      | |||| | ||| || | : | : | || | || | || | || | || | || | || |
Db    3 ACAATGCTCTTTTAGCCCTGCCATCTCGRGATGCCTTAAGTGTTTAAACAGCTCAGTGAA 61

RESULT 3
US-09-795-668-1349
; Sequence 1349, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795.668
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515.716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1349
; LENGTH: 61
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-668-1349

Query Match          0.2%; Score 23.4; DB 10; Length 61;
Best Local Similarity 61.0%; Pred. No. 1.2e+04;
Matches 36; Conservative 1; Mismatches 22; Indels 0; Gaps 0;

QY   2233 AAATGGACCTTTAGGGAAGATCTTTCAGGATAGCTGAAATGTTCAGAACCCCAAGTGAA 2291
      | |||| | ||| || | : | : | || | || | || | || | || | || | || |
Db    3 ACAATGCTCTTTTAGCCCTGCCATCTCGRGATGCCTTAAGTGTTTAAACAGCTCAGTGAA 61

RESULT 4
US-09-795-686-1349
; Sequence 1349, Application US/09795686
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795.686
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515.715
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
```



```

; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 920010.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 514:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 47 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-263-959-514

Query Match 0.2%; Score 22.2; DB 10; Length 47;
Best Local Similarity 69.8%; Pred. No. 2.1e+04;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4956 CACACACATCATCGAGGAGACTACACACACACACACAGGCCA 4998
||||| ||||| ||||||||| ||||| |
Db 47 CACACACACATCTCACACACATACACACACACACACACACACA 5

RESULT 11
US-09-983-965-5096/c
; Sequence 5096, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagapan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5096
; LENGTH: 63
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 31-LIB34-084-Q1-E1-H11
US-09-983-965-5096

Query Match 0.2%; Score 22.2; DB 10; Length 63;
Best Local Similarity 61.0%; Pred. No. 2.5e+04;
Matches 36; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 7506 ACGGACATCAAGGGAGACTACGCAAAACACACACAGAGCCCAACAGGAGATAGTAAGACCA 7564
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60 ACGGACACAAAGGCAGAAAAAGAAAAAAGAGAGCCCAACAAAAAATAGAAA 2

RESULT 12
US-10-057-940-14/c
; Sequence 14, Application US/10057940
; Patent No. US20020168686A1
; GENERAL INFORMATION:
; APPLICANT: Pantoliiano, Michael W.
; APPLICANT: Salemme, F. Raymond
; APPLICANT: Carver, Jr., Theodore, E.
; TITLE OF INVENTION: High Throughput Method for Functionally Classifying Proteins
; TITLE OF INVENTION: Identified Using a Genomics Approach
; FILE REFERENCE: 1503.031002/JAG/JSO
; CURRENT APPLICATION NUMBER: US/10/057,940

```

[illegible]

```
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697

Query Match          0.2%: Score 21.8; DB 9; Length 47;
Best Local Similarity 78.8%; Pred. No. 2.7e+04;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3610 TTACTGCGAGCAAAAGACAGCTACAGACTCT 3642
   | ||| ||||| ||||| || ||| | |||||
Db  8 TGACGAGGAGCAAAAGCAACTATGGGCTCT 40

RESULT 14
US-09-978-697-378
; Sequence 378, Application US/09978697
; Patent No. US20020169284A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2630PlC27
; CURRENT APPLICATION NUMBER: US/09/978,697
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081952
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
```

PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084598
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085689
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697

Query Match 0.2%; Score 21.8; DB 9; Length 47;
Best Local Similarity 78.8%; Pred. No. 2.7e+04;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3610 TTACCTGGCAGCAAAAGACAGCTACAGACTCCT 3642
| | | | | | | | | | | | | | | | | | | | |
Db 8 TGACCAGGAGCAAAAGGCAACTATGGCTCCT 40

RESULT 15
US-09-982-610-42
; Sequence 42, Application US/09982610
; Patent No. US20020146420A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; Bennett, Brian D.
; Goeddel, David
; Lee, James M.
; Matthews, William
; Tsai, Siao Ping
; Wood, William I.
; TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/982,610
; FILING DATE: 17-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/446,648
; FILING DATE: 1996-MAY-23
; APPLICATION NUMBER: 08/222616
; FILING DATE: 04-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P0821P3PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-982-610-42

Query Match 0.2%; Score 21.8; DB 10; Length 54;
Best Local Similarity 65.3%; Pred. No. 2.9e+04;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2799 GGAATAATTGTAATTAAGAAACGATGAAAGCAATGAAG 2847
| | | | | | | | | | | | | | | | | | | | |
Db 3 GGATATCATGGACTACAGGACGATGACAGAAGCTTGCCATGGAG 51

Search completed: December 3, 2002, 09:05:14
Job time : 260 secs

---



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: December 2, 2002, 23:47:05 ; Search time 7932 Seconds  
(without alignments)  
19940.132 Million cell updates/sec

Title: US-09-700-906A-1\_COPY\_197\_9962  
Perfect score: 9766  
Sequence: 1 atgtggccacgagacgct.....tcattaggacagtgaagata 9766

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 176248

Minimum DB seq length: 0  
Maximum DB seq length: 66

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

- 1: em\_estba:\*\*
- 2: em\_esthum:\*\*
- 3: em\_estin:\*\*
- 4: em\_estmu:\*\*
- 5: em\_estov:\*\*
- 6: em\_estpl:\*\*
- 7: em\_estro:\*\*
- 8: em\_htc:\*\*
- 9: gb\_est1:\*\*
- 10: gb\_est2:\*\*
- 11: gb\_htc:\*\*
- 12: gb\_est3:\*\*
- 13: gb\_est4:\*\*
- 14: gb\_est5:\*\*
- 15: em\_estfun:\*\*
- 16: em\_estom:\*\*
- 17: gb\_gss:\*\*
- 18: em\_gss\_hum:\*\*
- 19: em\_gss\_inv:\*\*
- 20: em\_gss\_pln:\*\*
- 21: em\_gss\_vrt:\*\*
- 22: em\_gss\_fun:\*\*
- 23: em\_gss\_mam:\*\*
- 24: em\_gss\_mus:\*\*
- 25: em\_gss\_Other:\*\*
- 26: em\_gss\_pro:\*\*
- 27: em\_gss\_rod:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match %	Description
c 1	36.8	0.4	52 14 T76979
	28.4	0.3	31 9 AA876462
	26.8	0.3	47 17 AZ456727
c 3	26.8	0.3	55 9 AU267450
	25.8	0.3	57 9 AJ443729
c 6	25.8	0.3	62 17 AZ777647

c 7	25.8	0.3	66 17 AZ780364
c 8	25.6	0.3	58 13 BI846606
c 9	25.6	0.3	61 17 CNS02MCM
c 10	25.4	0.3	62 9 A1125009
c 11	25.2	0.3	63 13 BI457211
c 12	25.2	0.3	66 9 A1887645
c 13	25	0.3	59 17 AZ375608
c 14	25	0.3	61 12 BF663117
c 15	24.8	0.3	66 9 AA215584
c 16	24.6	0.3	41 17 AZ822347
c 17	24.6	0.3	63 9 AL632972
c 18	24.6	0.3	64 17 AZ788410
c 19	24.4	0.2	61 17 CNS039XA
c 20	24.2	0.2	61 17 AZ953243
c 21	24.2	0.2	62 17 TAL05A10P
c 22	24.2	0.2	65 10 AW192956
c 23	24.2	0.2	65 13 BI845566
c 24	24	0.2	55 10 BE569086
c 25	24	0.2	60 14 B0548349
c 26	24	0.2	62 10 AV948136
c 27	24	0.2	62 10 AW633014
c 28	24	0.2	63 2 HSM001876
c 29	24	0.2	64 9 AA242896
c 30	24	0.2	65 9 AA590496
c 31	23.8	0.2	55 17 AZ788077
c 32	23.8	0.2	58 10 AV954391
c 33	23.8	0.2	66 9 AI950579
c 34	23.8	0.2	66 17 CNS06CZO
c 35	23.6	0.2	48 17 AZ477776
c 36	23.6	0.2	54 17 AZ314357
c 37	23.6	0.2	64 17 AZ815499
c 38	23.6	0.2	65 17 AZ770866
c 39	23.6	0.2	65 17 CNS03JF6
c 40	23.6	0.2	66 17 AZ373200
c 41	23.4	0.2	49 17 AZ659694
c 42	23.4	0.2	52 9 AA114400
c 43	23.4	0.2	54 9 AA053228
c 44	23.4	0.2	59 17 AZ589795
c 45	23.4	0.2	60 14 W52237

ALIGNMENTS

RESULT 1	T76979	52 bp	linear	EST 15-MAR-1995
T76979/c	YD72b10.s1	Soares fetal liver spleen	INFLS Homo sapiens	CDNA clone
LOCUS	IMAGE:113755 3'	similar to SP:B48666	B48666	CELL PROLIFERATION
DEFINITION	ANTIGEN KI-67, SHORT FORM -	;	mrna sequence.	
ACCESSION	T76979			
VERSION	T76979.1	GI:694182		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
	1 (bases 1 to 52)			
	Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman			
	, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,			
	Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston			
	, R., Williamson, A., Wohldmann, P. and Wilson, R.			
	The WashU-Merck EST Project			
	Unpublished (1995)			
TITLE	Other_ESTs: yd72b10.r1			
JOURNAL	Contact: Willson RK			
COMMENT	Washington University School of Medicine			
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108			
	Tel: 314 286 1800			
	Fax: 314 286 1810			
	Email: est@watson.wustl.edu			
	Insert Size: 2430			
	High quality sequence starts: 1			
	High quality sequence stops: 1			

c 1	36.8	0.4	52 14 T76979
c 2	28.4	0.3	31 9 AA876462
c 3	26.8	0.3	47 17 AZ456727
c 4	26.8	0.3	55 9 AU267450
c 5	25.8	0.3	57 9 AJ443729
c 6	25.8	0.3	62 17 AZ777647

Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Insert Length: 2430 Std Error: 0.00

```

Seq primer: -21ml3
High quality sequence stop: 1.
Location/Qualifiers
1. .52
/organism="Homo sapiens"
/db_xref="GDB:469372"
/db_xref="taxon:9606"
/clone="IMAGE:113755"
/clone_lib="Soares fetal liver spleen INFUS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' ACTCGAGCAATTAATTAAGATCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
3 a 21 c 13 g 13 t 2 others

```

BASE COUNT	ORIGIN
3 a	21 c 13 g 13 t 2 others

Query Match 0.4%; Score 36.8; DB 14; Length 52;  
Best Local Similarity 82.0%; Pred. No. 4.2e+03;  
Matches 41; Conservative 0; Mismatches 9; Indels

Oy 9485 ATGAGAGTCCCAGCTAAGGTGGCAGGAGAGCGGAGGGCAGAAGT 9534  
| | | | |  
Db 51 ACGAGTCCCCAGCCNAGGTGGCAGGAGAGCGGAGGGCAGAAGT 2

Db 51 ACGAGCTCCCCAGCCCNAGGTGGCAGAGGAGAGCGGAGGCGAGAAGAGT 2

RESULT 2  
AA876462 LOCUS  
DEFINITION  
AA876462 linear 31 bp mRNA EST 25-MAR-1998  
sw79h11.s1 NCI\_CGAP\_Pr12 Homo sapiens cDNA clone IMAGE:1252869  
similar to SW:K167\_HUMAN P45013 ANTIGEN KI-67. ;, mRNA sequence.  
ACCESSION  
AA876462  
VERSION  
AA876462.1 GI:2985539  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Robert Strausberg, Ph.D.

Email: [cgabps-remail@nih.gov](mailto:cgabps-remail@nih.gov)  
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,  
Rodrigo F. Chuquai, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

```
Trace considered overall poor quality
Seq primer: -40ml3 fwd. ET from Amerisham
High quality sequence stop: 1.
Location/Qualifiers
1. .31
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source
```

BASE COUNT	13 a	7 c	7 g	4 t
ORIGIN	NIH. average insert size 600 bp. Library made by D. Krizman, Non-directionally cloned. Size-selected on agarose gel, lesion of the bone. cDNA made by oligo-dT priming. /note="Vector: pAMP10; mRNA made from metastatic prostate /lab_host="DH10b" /tissue_type="metastatic prostate bone lesion" /sex="male" /clone_lib="NCI_CGAP_Pr12" /clone="IMAGE:1252869"			

BASE COUNT	13 a	7 c	7 g	4 t
ORIGIN				
Query Match	0.3%	Score 28.4;	DB 9;	Length 31;

Query Match	0.3%	Score 28.4;	DB 9;	Length 31;
Best Local Similarity	96.7%	Pred. No. 1.7e+05;		
Matches 29;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

Qy	2565	GACAAAACCTTCAGATACTGACACAGAGCC	2594
Dδ	1	GACAAAACCTTCAGATACTGACACAGAGGC	30

Db 1 GACAAACTTCAGATACTGAGACAGAGGC 30

RESULT 3					
AZ456727/c					
LOCUS	AZ456727	47 bp	DNA	linear	GSS 04-OCT-2000
DEFINITION	1M0259D17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0259D17 R, DNA sequence.				

ACCESSION	AZ430727
VERSION	AZ456727.1
KEYWORDS	GT:10614852
SOURCE	GSS.
ORGANISM	house mouse.
	Mus musculus
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 47)
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

**JOURNAL** Unpublished (2000)

**COMMENT** Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLCT, UT  
84112, USA

[illegible]

RESULT	5	57 bp	linear	EST
AJ443729	LOCUS	dkfz436	gallus cdna clone 37gllr1,	19-APR-2002
AJ443729	DEFINITION			
AJ443729	ACCESSION			
AJ443729.1	VERSION	GI:20210950		

**KEYWORDS**  
SOURCE chicken.  
**ORGANISM**  
Gallus gallus  
Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 57)  
Buerstedde,J.M.  
Gallus gallus bursal lymphocyte EST  
Unpublished (2002)  
**COMMENT**  
Cellular Immunology  
Heinrich-Pette-Institute  
Martinistr. 52, 20251 Hamburg, Germany  
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.

**FEATURES**  
source  
1..57  
/organism="Gallus gallus"  
/strain="CB"  
/db\_xref="taxon:9031"  
/clone\_lib="37gllrl"  
/clone\_lib="dkfz426"  
/tissue\_type="Bursa of Fabricius"  
/cell\_type="Bursal lymphocyte"  
/dev\_stage="2-3 weeks old"  
BASE COUNT 28 a 10 c 4 g 15 t  
ORIGIN

Query Match 0.3%; Score 25.8; DB 9; Length 57;  
Best Local Similarity 67.94; Pred. No. 6.4e+05;  
Matches 36; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2764 ATGAAGGAATAGAAACCTTTTGAGACATATAAGGAAAATTGAATTAAA 2816  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 2 ATAAGCACACAAAAACCTTTTAAGTTTACAAGGAATATTTCCTAAAA 54

**RESULT 6**  
AZ777647/c  
**LOCUS**  
2M0012N33F Mouse 10kb plasmid UGCIM library Mus musculus genomic  
clone UGC2NM0012N23 F, DNA sequence.  
AZ777647  
Accession  
Version AZ777647.1 GI:12906659  
Keywords GSS.  
Source house mouse.  
Organism Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 62)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
.M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLT, UT,  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0012 row: N column: 23  
Seq primer: CGTTGTAAACGACGGCAGT  
Class: plasmid ends  
High quality sequence stop: 62.  
Location/Qualifiers  
1..62  
/organism="Mus musculus"

**FEATURES**  
source

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUC2M0012N23"
/clone_lib="Mouse 10kb plasmid UUC1m library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      4 a   13 c   20 g   25 t
ORIGIN

Query Match      0.3%; Score 25.8; DB 17; Length 62;
Best Local Similarity 73.3%; Pred. NO. 6.5e+05;
Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 6075 CACACACACACAGAGACAGGAGGATGGAAGAGACATCAA 6119
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 58 CACACACACACAAAGGACAGACACTGAAGAAAGTCTCCCA 14

RESULT 7
LOCUS      A2780364
DEFINITION 2M0017J07R Mouse 10kb plasmid UUC1m library Mus musculus genomic
ACCESSION  A2780364
VERSION     A2780364.1 GI:12911951
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 66)
AUTHORS    Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
TITLE      Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL    Unpublished (2000)
COMMENT    Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0017 row: J column: 07
            Seq primer: CACACAGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 66.
            Location/Qualifiers
            1. .66
            /organism="Mus musculus"

```

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUC2M0017J07"
/clone_lib="Mouse 10kb plasmid UUC1m library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      2 a   16 c   2 g   46 t
ORIGIN

Query Match      0.3%; Score 25.8; DB 17; Length 66;
Best Local Similarity 63.9%; Pred. NO. 6.6e+05;
Matches 39; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 2789 AGACATATAAGGAAATATTGTAATTAAGAAACGATCAAGATGAAGCAATCAAGA 2848
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62 AAACAAAAAGAAATAAAGAAAGAAAAAGAGAGAAAGAAAGAAAGATGCAGG 3

QY 2849 G 2849
      |
DB 2 G 2

RESULT 8
LOCUS      BI846606/c
DEFINITION fQ80906.x1 Zebrafish neuronal Danio rerio cDNA clone 4886602 3',
            mRNA sequence.
ACCESSION  BI846606
VERSION     BI846606.1 GI:15959129
KEYWORDS   EST.
SOURCE     zebrafish.
ORGANISM   Danio rerio
REFERENCE  1 (bases 1 to 58)
AUTHORS    Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
            WashU Zebrafish EST Project 1998
            Unpublished (1998)
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrfish@watson.wustl.edu
            cDNA library constructed by S. Lin DNA Sequencing by: Washington
            University Genome Sequencing Center Clone distribution: the
            I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
            Seq primer: T7 from Gibco.

```

```

Query Match      0.3%; Score 25.6; DB 17; Length 61;
Best Local Similarity 63.8%; Pred. No. 7.1e+05;
Matches 37; Conservative 1; Mismatches 20; Indels 0; Gaps 0;

Qy 2789 AGACATATAAGCAAAATATTCGAATATAAAGAAAGGATGAAAGGATGAAGCAATGAA 2846
      ||||| || | - | - | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 59 ACACAAAAAAAWAAATAATGAAAAAAGAAAAAAGAAAAAAGAAAAAATAATAAAGAAA 2

RESULT 10
AIL125009/c
LOCUS      AIL125009                62 bp      mRNA      linear      EST 11-SEP-1998
DEFINITION aol3909.sl Barstead aorta HPLRB3 Homo sapiens cDNA clone
IMAGE:11726528 3', mRNA sequence.
ACCESSION  AIL125009
VERSION    AIL125009.1 GI:3593523
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 62)
AUTHORS   Hillier,L., Allten,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,f., Wyllie,f., Waterston,R. and Wilson,R.
TITLE     WashU-NCI human EST Project
JOURNAL   Unpublished (1997)
COMMENT   Contact: Wilton RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40ml3 fwd. Ef from Amersham.

FEATURES             source
     location/Qualifiers
         1..62
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:1726528"
             /clone_lib="Barstead aorta HPLRB3"
             /sex="male"
             /tissue_type="aorta"
             /dev_stage="adult, age 64"
             /lab_host="DH10B"
             /note="vector; pT73D-Pac (Pharmacia) with a modified
             polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
             was primed with a Not I oligo(dT) primer [5',
             TGTTACCAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTTTTTTTTTT
             3']; double-stranded cDNA was ligated to Eco RI adaptors
             [5', AATTCGGATCCCAAC 3' and 5' GTTGATCCG 3'], digested
             with Not I and cloned into the Not I and Eco RI sites of
             the modified pT73 vector. Library constructed by Bob
             Barstead."
     BASE COUNT      8 a      6 c      4 g      44 t
     ORIGIN
Query Match      0.3%; Score 25.4; DB 9; Length 62;
Best Local Similarity 64.4%; Pred. No. 7.8e+05;
Matches 38; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 9307 CAGCAATAACTGAGGTCTTTTGTATTAGCAGAAAGAAATAGAAATAAACAGAAATGAAAA 9365
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 59 CAGTAAAAATCTGAATGTTGTCAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1

RESULT 11
BI457211
LOCUS      BI457211                63 bp      mRNA      linear      EST 21-AUG-2001
DEFINITION 603185478F1 NIH MGC 42 Homo sapiens cDNA clone IMAGE:5258275 5'.

```

```

ACCESSION B1457211
VERSION B1457211.1 GI:15247867
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 63) .
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1869 row: 1 column: 20
High quality sequence stop: 63.
FEATURES             Location/Qualifiers
     source            1..63
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone="IMAGE:5258275"
     /clone_lib="NIH_MGC_42"
     /tissue_type="epitheloid carcinoma cell line"
     /lab_host="DH10B (phage-resistant)"
     /note="Organ: pancreas; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene), and Superscript II RT (Life Technologies)."
Note: this is a NIH_MGC Library. |"
BASE COUNT      22 a   18 c   20 g   3 t
ORIGIN
Query Match          0.3%; Score 25.2; DB 13; Length 63;
Best Local Similarity 78.9%; Pred. No. 8.6e+05;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 9055 GAGGAATTCGTGGAGCGCTGCCAGCCCAAGAAGCA 9092
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GAGAAATCAGCAGCAGCGTGTCAGCGCCCAAGAAGCA 38

RESULT 12
AI887645/C
LOCUS wml60ll.xl NCI_CGAP_Ut4 Homo sapiens CDNA clone IMAGE:2436117 3',
DEFINITION mRNA sequence.
ACCESSION AI887645
VERSION AI887645.1 GI:5592809
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 66)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@remail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA library preparation: Life Technologies, Inc.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -400P from Glbco.
FEATURES             Location/Qualifiers
     source            1..66
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone="IMAGE:2436117"
     /clone_lib="NCI_CGAP_Ut4"
     /tissue_type="serous papillary carcinoma, high grade, 2
pooled tumors"
     /lab_host="DH10B"
     /note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: SaltI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.48 kb. Life Technologies catalog #:
11542-016"
BASE COUNT      13 a    2 c    5 g    46 t
ORIGIN
Query Match          0.3%; Score 25.2; DB 9; Length 66;
Best Local Similarity 62.9%; Pred. No. 8.7e+05;
Matches 39; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 9307 CAGCAAATAACTGAGGCTTTGTATTAGCAGAAGAAATACAATAAACAGAAATGAAAAG 9366
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 CAGTAACCCCTAATTTTTTTTGTAAAAAANAANAANAANAANAANAANAANAANA 6

QY 9367 AA 9368
||
Db 5 AA 4

RESULT 13
AZ375608/C
LOCUS AZ375608 Mouse 10kb plasmid UUGCLM library Mus musculus genomic
DEFINITION linear DNA sequence.
ACCESSION AZ375608.1 GI:10489308
VERSION AZ375608
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 59)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0129 row: W column: 04
Seq primer: CGTTGTAACACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 59.
FEATURES             Location/Qualifiers
     source            1..59
     /organism="Mus musculus"
     /strain="C57BL/6J"
     /db_xref="taxon:10090"

```

following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library.

6 a	19 c	17 q	19 q
-----	------	------	------

BASE COUNT

KEYWORDS	ORGANISM	EST.	SOURCE
	Homo sapiens		human.
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 66)		

REFERENCE  
1 (bases 1 to 66)  
AUTHORS  
TITLE  
JOURNAL

unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapsb-f@mail.nih.gov  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -41ml3 fwd. ET from AmerSham  
High quality sequence stop: 50.  
Location/Qualifiers  
source 1. .66

germinal center B cells by flow sorting (CD20<sup>+</sup>, IgD<sup>-</sup>), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer (5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3' [1]). Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

```

BASE COUNT      7 a      1 c      4 g      54 t
ORIGIN
Query Match      0.3%; Score 24.8; DB 9; Length 66;
Best Local Similarity 63.3%; Pred. No. 1e+06;
Matches 38; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 2801 AAAATATTGTAATTAAAGAAAAACGATGAAAGATGGAAGCAATGAAGAGATCAAGAACCTT 2860
|||||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 60 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAACTT 1
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Search completed: December 3, 2002, 09:01:01  
Job time : 7967 secs

**THIS PAGE BLANK (USPTO)**